REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	12 Aug 95	Final 15 Jun 92 - 14 Jun 95

4. TITLE AND SUBTITLE

Automated Target Tracking and Recognition Using Jump-Diffusion Processes

5. FUNDING NUMBERS

DAAL03-92-G-0141

6. AUTHOR(S)

Michael I. Miller
Anuj Srivastava

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Washington University
Department of Electrical Engineering
St. Louis, MO 63130

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 29349.4-MA-SDI

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (

This report presents our work, supported under the research grant ARO DAAL03-92-G-0141, on the development of an algorithm for generating the conditional mean estimates of functions of target positions, orientation and type in recognition and tracking of an unknown number of targets and target types. Taking a Bayesian approach a posterior measure is defined on the tracking/target parameter space by combining the narrowband sensor array manifold model with a high resolution imaging model, and a prior based on airplane dynamics. The Newtonian force equations governing rigid body dynamics are utilized to form the prior density on airplane motion. The conditional mean estimates are generated using a random sampling algorithm based on *Jump-Diffusion* processes, [1], for empirically generating MMSE estimates of functions of these random target positions, orientations and type under the posterior measure. Results are presented on target tracking and identification from an implementation of the algorithm on a networked Silicon Graphics and DECmpp/MasPar parallel machines.

14. SUBJECT TERMS

Multi-Target Tracking/Recognition, Random Sampling, Jump-Diffusion

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

AUTOMATED TARGET TRACKING AND RECOGNITION USING JUMP-DIFFUSION PROCESSES

Authors:

MICHAEL I. MILLER & ANUJ SRIVASTAVA

DATE: August 12th, 1995

FINAL TECHNICAL REPORT

RESEARCH AGREEMENT NUMBER: DAAL03-92-G-0141

Contents

List of Figures

Abstract

This report presents our work, supported under the research grant ARO DAAL03-92-G-0141, on the development of an algorithm for generating the conditional mean estimates of functions of target positions, orientation and type in recognition and tracking of an unknown number of targets and target types. Taking a Bayesian approach a posterior measure is defined on the tracking/target parameter space by combining the narrowband sensor array manifold model with a high resolution imaging model, and a prior based on airplane dynamics. The Newtonian force equations governing rigid body dynamics are utilized to form the prior density on airplane motion. The conditional mean estimates are generated using a random sampling algorithm based on *Jump-Diffusion* processes, [1], for empirically generating MMSE estimates of functions of these random target positions, orientations and type under the posterior measure. Results are presented on target tracking and identification from an implementation of the algorithm on a networked Silicon Graphics and DECmpp/MasPar parallel machines.

Here we present the final report on the research conducted under the research grant DAAL03-92-G-0141 from Army Research Office. This research was conducted at Electronic Signals and Systems Research Laboratory, Washington University, in collaboration with Prof Ulf Grenander at Division of Applied Mathematics, Brown University. Following are the publications and presentations which resulted from this project.

Publications:

1. M. I. Miller, A. Srivastava and U. Grenander, *Conditional-Expectation Estimation Via Jump-Diffusion Processes in Multiple Target Tracking/Recognition*, IEEE Transactions on Signal Processing, to appear November, 1995.

2. A. Srivastava, M. I. Miller and U. Grenander, *Multiple Target Direction of Arrival Tracking*, IEEE Transactions on Signal Processing, vol. 43, number 5, May, 1995, pages 1282-85.

3. Anuj Srivastava, *Stochastic Processes on Lie Groups for Automated Target Tracking & Recognition*, Dissertation Proposal, Washington University, April, 1995.

4. U. Grenander and M. I. Miller, *Representations of Knowledge in Complex Systems*, Journal of the Royal Statistical Society, 56(3), 1994.

Presentations:

1. M. A. Foltz, A. Srivastava, M. I. Miller and U. Grenander, *Detection of Multiple Airborne Targets from Multi-Sensor Data*, SPIE Conference, San Diego, CA., 1995.

2. A. Srivastava and M. I. Miller, *Bayesian Framework for Automated Detection, Tracking and Recognition*, ONR/NAWC Workshop on *Acquisition and Tracking of Maneuvering Targets from Image Sequence Data* at China Lake, CA, on May 16-17th, 1995.

3. A. Srivastava, M. I. Miller and U. Grenander, *Lie Group Parameterization for Dynamics Based Prior in ATR*, Proc. Sixth Digital Signal Processing Workshop, Yosemite, Ca., October 2-5, 1994.

4. A. Srivastava, R. S. Teichman and M. I. Miller, *Target Tracking and Recognition Using Jump-Diffusion processes*, Army Research Office's *Eleventh Army Conference on Applied Mathematics and Computing*, Carnegie Melon University, Pittsburgh, PA, June, 1993.

5. R. S. Teichman, A. Srivastava and M. I. Miller, *Automated Tracking and Target Recognition Algorithms Using Jump-Diffusion Processes*, Rome Laboratory's *3rd Annual 1993 IEEE Dual Use Technologies & Application Conference* at Utica, NY, on May 24-27, 1993.

6. A. Srivastava, R. S. Teichman, M. I. Miller, D. L. Snyder and J. A. O'Sullivan, *Jump-Diffusion Based Sampling Algorithm For Target Tracking and Recognition*,Proc. 27th Annual Asilomar Conference on Signals, Systems & Computers, Pacific Grove, Ca., November, 1993.

7. M. I. Miller, R. S. Teichman, A. Srivastava, J. A. O'Sullivan and D. L. Snyder, *Jump-Diffusion Processes for Automated Tracking-Target Recognition*, 1993 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, Maryland, 1993.

8. A. Srivastava, N. Cutaia, M. I. Miller, J. A. O'Sullivan and D. L. Snyder, *Multi-target narrowband direction finding and tracking using motion dynamics*, Proc. 30th Annual Allerton Conference on Communication, Control, and Computing, 1992.

The first two publications are also presented as appendices (C & D) to this report.

1 Introduction

Our work under the contract ARO DAAL03-92-G-0141 focuses on automated tracking and recognition of objects in remotely sensed complex dynamically changing scenes. Grenander's global shape models are used herein, extended to parametric representations of arbitrary and unknown model order, in which typical shape is represented via templates, with variability represented via transformation groups applied to the templates. The types of variability associated with the classical geometry are accommodated via the *Euclidean groups* involving both the rigid motions of translation and rotation. Since the objects are under dynamic motion, *the parameter spaces involves Cartesian products of these similarity groups.*

The second fundamental type of variability is associated with the model order (parametric dimension) and model type (recognition). In any scene there may be variable numbers of and different kinds of targets existing in the scenes for varying periods of time, implying the target number and therefore parametric dimension are unknown apriori. *Hence, the inference or hypothesis space becomes a search across countable disconnected unions of these Cartesian product groups*, with the model order and model type a variable to be inferred. We take a Bayesian approach, i.e. we define a prior distribution supported on this countable union of spaces, from which the posterior distribution is constructed. The parametric representation of the target scene is selected to correspond to *conditional expectations* under this posterior.

As we are particularly interested in non-cooperative moving targets, the algorithms are made robust to motion by incorporation of knowledge about motion dynamics into the prior distribution. The Newtonian force equations, a system of differential equations governing the motion of targets are used to induce the prior. These differential equations are parameterized

by the target and or sensor type, and its orientation motion described by rotations in the 3-dimensional torus group. It is the introduction of these Newtonian force equations which makes tracking and recognition inseparable, since the equations of motion are explicitly parameterized by the sequence of airplane orientations. This provides the significant link between tracking algorithms based on data from narrowband sensors arrays in which the target is unresolved in the data (effectively a point), and high resolution information perhaps provided by a second sensor preserving the orientation information from which target recognition is performed. In part, it is this fundamental link which has motivated us to solve the tracking/recognition problem in a single consistent estimation framework in which the inference proceeds via the fusion of multi-sensor data: in our case, a *narrowband sensor array output* and *high-resolution images.*

Now automated target tracking and recognition are well known problems in the signal processing and control system's literature, with a great deal of published work on multiple target tracking posed as state estimation problems [5, 6, 7]. In such approaches Kalman filter based techniques are emphasized, with linear descriptions of state playing a fundamental role. For situations in which the observed data are non-linear in target parameters the use of the extended Kalman filter has been proposed corresponding to linear approximations which prove valid for particular scenarios. There also now exists a substantial body of important work in tracking the directions of arriving signals from multiple moving sources recorded via sensor arrays [8, 9, 10]. In such sensor array based approaches the non-linear relationship between the parameters of motion and the sensor data are addressed directly, the linear Kalman filter state equations for tracking guiding or providing initial conditions for the gradient based estimators generated from the likelihood. In these non-linear data models, several variations of the gradient based techniques are used to solve the problem in mostly maximum-likelihood settings. However, the majority of researchers utilize simplifying assumptions which are not always valid in a general tracking scenario. For example, targets may be assumed stationary between sample times with multiple ($\sim$ 100) snapshots at each sample time, whereas, in general, for a moving target, each data sample reflects a new position. Also, though researchers base their models on simplified versions of target dynamics for the tracking scenario, mostly constant velocity - constant acceleration state constraint equations have been used because of their linear nature. These restricted motions are partly due to assumptions required for Kalman updating, but perhaps more fundamentally due to the separation of the tracking and recognition problems. The more informative priors used in this report require high resolution recognition as the priors are coupled to the target type and its orientations. In part, this is one of the major results of this work.

1.1 Random Sampling Methodology

Concerning the generation of conditional expectations, except under the most simplifying set of assumptions, the posterior distribution will be highly nonlinear in the parameters of hypothesis space, thus, precluding the direct closed form analytic generation of conditional expectations. Towards this end we have taken advantage of the explosion which has occurred over the past 10 years in the statistics community on the introduction of random sampling methods for the empirical generation of estimates from complicated distributions; see for example the reviews [2, 3]. Motivated by such approaches, we have previously described a new family of random sampling algorithms [4, 1] for generating conditional expectations in such disconnected hypothesis spaces. The random samples are generated via the direct simulation of a Markov process whose state moves through the hypothesis space with the *ergodic property* that the transition distribution of the Markov process converges to the posterior distribution. This allows for the empirical generation of conditional expectations under the posterior. To accommodate the connected and disconnected nature of the state spaces, the Markov process is forced to satisfy *jump-diffusion dynamics.* i.e. through the connected parts of the parameter space (Lie manifolds) the algorithm searches continuously, with sample paths corresponding to solutions of standard diffusion equations; across the disconnected parts of parameter space the jump process determines the dynamics. The infinitesimal properties of these jump-diffusion processes are selected so that various sample statistics converge to their expectation under the posterior.

The original motivation for introducing jump-diffusions in [4, 1] is to accommodate the very different continuous and discrete components of the object discovery process. Given a conformation associated with a target type, or group of targets, the problem is to identify the orientation and translation parameters accommodating the variability manifest in the viewing of each object type. For this, the parameter space is sampled using diffusion search in which the state vector winds continuously through the similarities following gradients of the posterior. The second distinct part of the sampling process corresponds to the target type and number deduction during which the target types are being discovered, with some subset of the scene only partially "recognized" at any particular time during the process. The second type of change in parameter space are associated with a set of non-continuous transformations of the scene controlled by the jump process. A jump in hypothesis space corresponds to (i) jumping between different object types, (ii) hypothesizing a new object in the scene or a "change of mind" via the deletion of an object in the scene, or (iii) the merging or splitting of tracks and objects. The jump intensities are governed by the posterior density, with the process visiting configurations of higher probability for longer exponential times, and the diffusion equation governing the dynamics between jumps. It is the fundamental difference between diffusions (almost surely continuous sample paths) and jump processes (making large moves in parameter space in small

time) which allows us to explore the very different connected and non-connected nature of hypothesis space.

Under our research contract we have developed a random sampling based solution for generating minimum mean squared error estimates of the state variables for tracking and recognition problems in a general setting. We assume data from a narrowband sensor array providing azimuth-elevation data for object tracking, and optical or radar imagers providing detailed information about the target-type and orientation. The goal is to track and recognize the unknown number of non-cooperative sources.

1.2 Results and Contribution

We have presented a new random sampling algorithm for estimating the characteristics of moving signal sources. The method of estimation is to derive a single posterior distribution over the space $\mathcal{X}_t$ and then sample it via a Markov process $X(s)$ which satisfies jump-diffusion dynamics. This approach solves the full Bayesian problem as, in theory, no approximation is necessary. It is based on the work of Grenander and Miller [4, 1, 11] who have described a new class of sampling algorithms for a wide variety of applications including image analysis, crystallography and stochastic language models. These algorithms involve stochastic search over well defined parameter spaces following the Bayesian measures on these spaces.

An implementation of this algorithm for estimating a single track scene is presented. The algorithm was jointly implemented using a Silicon Graphics workstation for data generation and visualization, and a massively parallel 4096 processor SIMD DECmpp machine for implementing the tracking-recognition algorithm. It includes generating a parameterized target path using the Silicon Graphics' flight simulator which forms the true configuration. Using this path the simulated data are generated for both the tracking and imaging sensors. The generated data are then used in the estimation process to obtain the *MMSE* estimates of the actual flight.

As shown in the figure, the data collected from the sensors, observing diverse aspects of the scene, is fed into a unifying multi-sensor fusion scheme for joint solution to target detection, tracking and recognition problems. Following are the salient features of our ATR system resulting in better scene understanding and superior results.

- Incorporating multiple sensor outputs in unified, simultaneous estimation.
- Utilizing temporal processing, in form of powerful target dynamics, to compliment and suppliment data driven inferences.
- Heirarchial image understanding via *global shape models* for target localization and identification.
- Adapting algorithm to unknown & varying model order.

- Efficient implementation in a parallel processing environment.
- Framework for sequential estimation to handle incoming data flow.

Section 2 lists the set of parameters completely describing an observed scene and outlines the estimation problem. The posterior distribution on the parameter space is derived in section 3 by describing the dynamics based prior and observed data likelihoods. An estimation algorithm based on jump-diffusion processes is presented in section 4 along with the important theoretical results supporting the algorithm. Section 5 describes the use of a jump-diffusion algorithm for estimation of a single track configuration. It involves the implementation across machines such as the Silicon Graphics workstation and the DECmpp 12000 SIMD machine connected through data transfer on a high speed network.

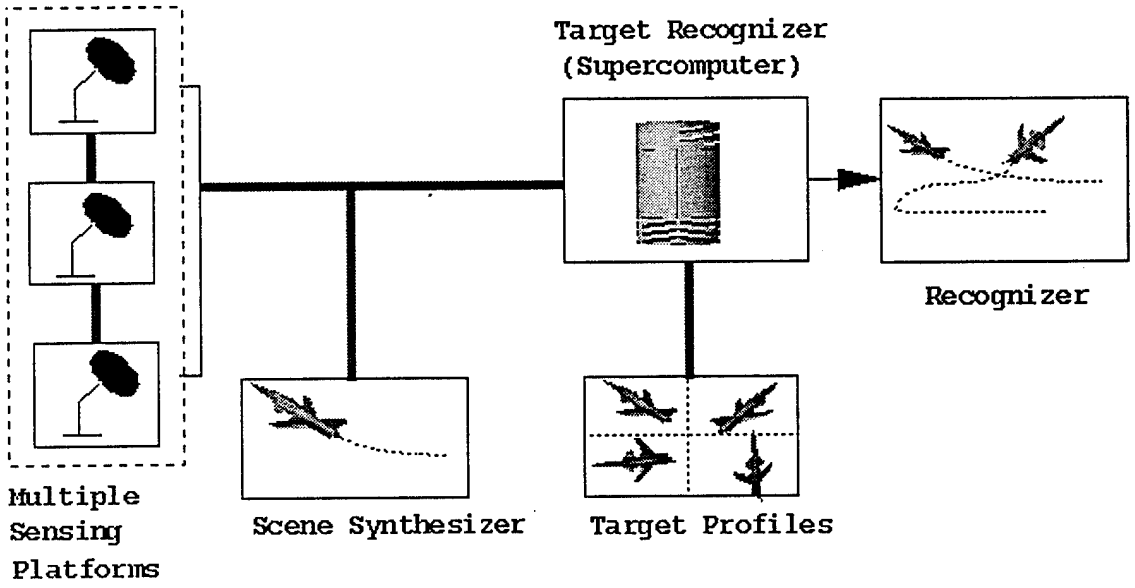


Figure 1: Overview of the automated target tracking and recognition system.

2 Scene Parameterization

2.1 Parameter Set

We use the global shape models and pattern theoretic approach introduced by Grenander [18, 19] to analyze complex scenes. As the basic building blocks of the hypotheses we define a subset of generators $\mathcal{G}^o$, which contains each target type $a \in \mathcal{A}$ placed at the origin of the inertial reference frame at a fixed orientation and unit scale. The fundamental variability in target spaces is accommodated by applying the transformations $T(\vec{\phi}), T(\vec{p}), T(s)$ to the templates $g^o \in \mathcal{G}^o$ according to

$$T(\vec{\phi}) : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi_1 & sin\phi_1 \\ 0 & -sin\phi_1 & cos\phi_1 \end{bmatrix} \begin{bmatrix} cos\phi_2 & 0 & sin\phi_2 \\ 0 & 1 & 0 \\ -sin\phi_2 & 0 & cos\phi_2 \end{bmatrix} \times \begin{bmatrix} cos\phi_3 & sin\phi_3 & 0 \\ -sin\phi_3 & cos\phi_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (1)$$

$$T(\vec{p}) : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 + p_1 \\ x_2 + p_2 \\ x_3 + p_3 \end{bmatrix}, \quad (2)$$

$$T(s) : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad (3)$$

where $\vec{\phi}$ is the triple of rotation angles (pitch, roll and yaw), $\vec{p}$ is the translation vector, and s is the scale parameter. These parameterized transformations operate on the templates from $\mathcal{G}^o$ generating the full set of elements $\mathcal{G}$. The observed scene at any time is modeled as a set of generators $G = \{g(1), g(2), .., g(M)\}$, each generator $g(m) \in \mathcal{G}$. Then, $\{a, \vec{\phi}, \vec{p}, s\}$ parameterize the representation of all possible targets generated from these transformations. Figure 2 shows one of the 3-D ideal targets used for all the simulations presented here. The left panel shows a rendering of the target generator $g^o \in \mathcal{G}^o$ at the origin, the right panel showing the result of applying one of the rotation transformations resulting in $g \in \mathcal{G}$.

2.2 Parametric Space

The set parameterizing the Bayes posterior becomes the set of parameters specifying the similarity transformations, as well as the airplane type. Define the space containing orientations $\vec{\phi}$ as the three dimensional torus $\mathcal{M}(3) \equiv [0, 2\pi]^3$ with $0, 2\pi$ identified. The position vector $\vec{p}$ belongs

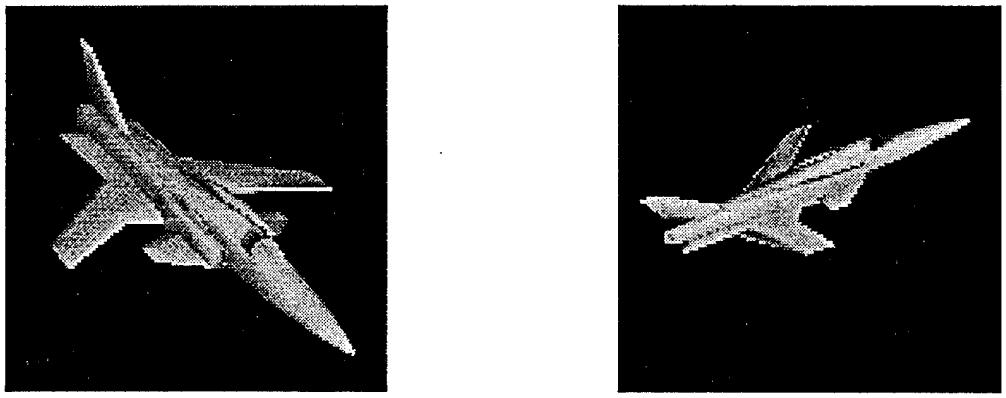


Figure 2: 3-D target generator $g \in \mathcal{G}^o$ at the origin (left panel) and after applying a rotation transformation (right panel).

to $\Re^3$ with the scale parameter belonging to $\Re_+$. Then associated with each target or generator $g \in \mathcal{G}$ at any time τ is a parameter set $x(\tau) = \{a(\tau), \vec{p}(\tau), \vec{\phi}(\tau), s(\tau)\} \in \mathcal{M}(3) \times \Re^3 \times \mathcal{A} \times \Re_+$, where $|\mathcal{A}| = |\mathcal{G}^o|$ the number of different target types.

A pattern will be constructed for the representation of the multiple track scenes with varying track lengths. We are interested in tracking-recognition in non-cooperative environments in which the m^{th} object appears and disappears at random times $t_1^{(m)}, t_1^{(m)} + t^{(m)}$ with its stay given by the interval $\mathcal{T}^{(m)} = [t_1^{(m)}, t^{(m)} + t_1^{(m)}]$. Clearly, $t_0 \leq t_1^{(m)} \leq t^{(m)} + t_1^{(m)} \leq t$, for the observation interval $[t_0, t]$. Define $x^{(m)}(\tau)$ to be the set of parameters associated with the target m at time τ given by $\{\vec{p}^{(m)}(\tau)\ , \vec{\phi}^{(m)}(\tau)\ , a^{(m)}(\tau)\ , s^{(m)}(\tau)\}$. For the observation period $[t_0, t]$ the parameter vector associated with the complete m^{th} track given $\mathcal{T}^{(m)}$ is

$$\{x^{(m)}(\tau) : \tau \in \mathcal{T}^{(m)}\} \in \left(\mathcal{M}(3) \times \Re^3 \times \Re_+ \times \mathcal{A}\right)^{\mathcal{T}^{(m)}} .$$

In real situations with discrete observations, the tracks get discretized to observation times $1, 2, \ldots . t \in \aleph$ (assume $t_0 = 1$ for simplicity). We denote the discrete parameters with the same symbols except now they belong to discrete sets, i.e. $t_1^{(m)}, t^{(m)}, t \in \aleph$ and $\mathcal{T}^{(m)} = \{t_1^{(m)} + 1, ..., t_1^{(m)} + t^{(m)}\}$. Hence the parameter set associated with the discretized track $\{x^{(m)}(k); k \in \mathcal{T}^{(m)}\}$ given $\mathcal{T}^{(m)}$ is an element of $(\mathcal{M}(3) \times \Re^3 \times \Re_+ \times \mathcal{A})^{t^{(m)}}$. Since the dwell time of the target, given by $\mathcal{T}^{(m)}$, is unknown a-priori its associated parameter set is an element of

$$\bigcup_{t^{(m)}=1}^{t} \left(\mathcal{M}(3) \times \Re^3 \times \Re_+ \times \mathcal{A}\right)^{t^{(m)}} \times \aleph .$$

The parameter vector for an M-track scene becomes the collection of each of the single track

parameter sets, element of $\mathcal{X}_t(M)$ according to

$$\begin{aligned}\vec{x}_t(M) &= \bigcup_{m=1}^{M} \{x^{(m)}(k) : k \in \mathcal{T}^{(m)}\} \\ &\in \mathcal{X}_t(M) \equiv \prod_{m=1}^{M} \left(\bigcup_{t^{(m)}=1}^{t} \left(\mathcal{M}(3) \times \Re^3 \times \Re_+ \times \mathcal{A}\right)^{t^{(m)}} \times \aleph \right) .\end{aligned}$$

For later convenience we also define $\vec{p}_t(M)$ to be the vector having elements the position components for all tracks and $\vec{\phi}_t(M)$ to be the vector containing orientation components of all tracks. Since M is unknown we define the complete configuration space $\mathcal{X}_t$ over which the estimation is performed as

$$\mathcal{X}_t = \bigcup_{M=0}^{\infty} \mathcal{X}_t(M) .$$

The estimation problem is to estimate the individual configurations as well as the number M. In this report, only rigid transformations are used with $s = 1$.

3 Bayesian Posterior

We take a Bayesian approach for solving the estimation problem by defining a posterior probability on the parameter space $\mathcal{X}_t$. As the posterior distribution is proportional to the product of the prior distribution and the observed data likelihood we first derive a prior measure on the parameter space followed by a model for the data generation which determines the likelihood term.

The prior measure encodes our *a-priori* information about the parameters to be estimated. In particular this knowledge can come from, say, the airplane dynamics, or some previous knowledge of target type and number of targets. Under this contract we have developed a system to utilize the set of Newton's second law based equations governing airplane motion to generate a prior distribution on the airplane paths. The prior on the orientation parameters is based on the von-Mises density. There are two types of data sets used here: the tracking data collected by a cross-array of isotropic sensors and the imaging data generated by optical sensing radar. The likelihood of the tracking data follows from the standard narrowband signal model, first proposed by Schmidt[12], whereas the imaging data are simply given by the far field projection of the actual 3-D object.

It needs to be emphasized that in real time estimation problems like these, where the data set is augmented at every observation time, the posterior distribution changes with time t, and is an explicit function of t denoted by $\pi_t(\cdot)$, $t \in \aleph$. Therefore, in this Bayesian approach the estimates are generated at any given time conditioned on the data accumulated up to that time.

3.1 Prior Density on Parameter Space $\mathcal{X}_t$

The formulation of the prior measure on the airplane positions $\vec{p}(s)$ is based on the equations of motion governing the airplane's flight.

3.1.1 Analysis of Airplane Motion

First, we derive the prior for a single target case by considering its underlying continuous motion. This derivation follows the description in [15], where the equations describing target dynamics are utilized to form a prior measure on the airplane positions $\vec{p}(s)$. These dynamics are easily expressed using the target velocities projected along the body-fixed axes, called the body-frame velocities $\vec{v}(s)$, as shown in Figure 3. Since the tracking array responds to the inertial positions of the target, we use the standard transformation to relate body-frame velocities with inertial frame positions given by

$$\vec{p}(s) = \int_{t_0}^{s} \Psi(\tau)\vec{v}(\tau)d\tau + \vec{p}(t_0), \tag{4}$$

where $\vec{p}(t_0)$ is assumed known and $\Psi(\tau)$ is a function of the Euler angles $\vec{\phi}(\tau)$ given by

$$\Psi(\tau) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi_1(\tau) & sin\phi_1(\tau) \\ 0 & -sin\phi_1(\tau) & cos\phi_1(\tau) \end{bmatrix} \begin{bmatrix} cos\phi_2(\tau) & 0 & sin\phi_2(\tau) \\ 0 & 1 & 0 \\ -sin\phi_2(\tau) & 0 & cos\phi_2(\tau) \end{bmatrix}$$

$$\times \begin{bmatrix} cos\phi_3(\tau) & sin\phi_3(\tau) & 0 \\ -sin\phi_3(\tau) & cos\phi_3(\tau) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which is same as the rotation matrix in Eqn 1. In general, $\Psi(\tau)$ converts any vector in the body frame of reference to the corresponding vector in the inertial frame. The rotation dynamics are described in terms of the angular velocities $\vec{q}(s)$ which are the known functions of the Euler angles $\phi(s)$ and their rates of change $\dot{\vec{\phi}}(s)$, given by ([20]),

$$\begin{aligned} q_1 &= \dot{\phi_1} - \dot{\phi_3} sin(\phi_2), \\ q_2 &= \dot{\phi_2} cos(\phi_1) + \dot{\phi_3} cos(\phi_2) sin(\phi_1), \\ q_3 &= -\dot{\phi_2} sin(\phi_1) + \dot{\phi_3} cos(\phi_2) cos(\phi_1). \end{aligned} \tag{5}$$

Following the rigid body analysis in [15], we neglect the earth's curvature, motion and wind effects. Then, the linear velocities $\vec{v}(s)$ and the angular velocities $\vec{q}(s)$ satisfy the following set of differential equations,

$$\begin{aligned} \dot{v}_1(s) - q_3(s)v_2(s) + q_2(s)v_3(s) &= f_1(s), \\ \dot{v}_2(s) + q_3(s)v_1(s) - q_1(s)v_3(s) &= f_2(s), \\ \dot{v}_3(s) - q_2(s)v_1(s) + q_1(s)v_2(s) &= f_3(s), \\ I_1\dot{q}_1(s) - (I_2 - I_3)q_2(s)q_3(s) &= \Gamma_1(s), \\ I_2\dot{q}_2(s) - (I_3 - I_1)q_1(s)q_3(s) &= \Gamma_2(s), \\ I_3\dot{q}_3(s) - (I_1 - I_2)q_2(s)q_1(s) &= \Gamma_3(s), \end{aligned}$$

where $\vec{f}(s) = [f_1(s)\ f_2(s)\ f_3(s)]$ is the vector of applied translational forces, $\vec{I} = [I_1\ I_2\ I_3]$ is the vector of rotational inertias, and $\vec{\Gamma}(s) = [\Gamma_1(s)\ \Gamma_2(s)\ \Gamma_3(s)]$ is the vector of applied torques. The first three equations describe the airplane's translational motion while the next three describe its rotational motion. We propose to use these equations to derive prior densities on the translational and rotational motions. In the work performed under this contract, we use only the first three equations for deriving priors on the positions using simple Markov priors for the rotational parameters instead of the last three equations. In vector form the first three

equations become

$$\dot{\vec{v}}(s) + A_1(\vec{\phi}(s), \dot{\vec{\phi}}(s))\vec{v}(s) = \vec{f}(s) \tag{6}$$

where

$$A_1(\vec{\phi}(s), \dot{\vec{\phi}}(s)) = \begin{bmatrix} 0 & -q_3(s) & q_2(s) \\ q_3(s) & 0 & -q_1(s) \\ -q_2(s) & q_1(s) & 0 \end{bmatrix},$$

the angular velocities $\vec{q}(s)$ being determined by the Euler angles $\vec{\phi}(s)$ and their rates of change. Since the eigen values of $A_1(\vec{\phi}(s), \dot{\vec{\phi}}(s))$ lie on the imaginary axis, this system is not stable. Following standard treatments (see [20] for example), we add a term $G\vec{v}(s)$ to the force vector, which provides a linear stabilizing feedback to the system. A simple diagonal gain matrix G is used, with the resulting system equation given by

$$\dot{\vec{v}}(s) + A(\vec{\phi}(s), \dot{\vec{\phi}}(s))\vec{v}(s) = \vec{f}(s), \tag{7}$$

where $A(\vec{\phi}(s), \dot{\vec{\phi}}(s)) = A_1(\vec{\phi}(s), \dot{\vec{\phi}}(s)) - G$. This linear vector differential equation is characterized by the time-varying parameter matrix $A(\vec{\phi}(s), \dot{\vec{\phi}}(s))$ which depends on the rotational motion of the airplane. The prior is induced following the approach in Amit et al. [21] and used in [15] by assuming the forcing function to be a white process with fixed spectral density σ_0^2. As the equations are linear in velocity vector this induces a Gaussian prior on $\vec{v}(s)$ conditioned on the Euler angles with the covariance determined by solving the differential equation 7.

Define the state transition matrix $\Phi(\tau, \cdot)$ as the unique solution of the matrix differential equation

$$\frac{dM(s)}{ds} = -A(\phi(s), \dot{\phi}(s))M(s)\ , \quad M(\tau) = I\ , \tag{8}$$

then the covariance of the body frame velocity process becomes

$$\mathcal{K}_v(s_1, s_2) = \sigma \int_{t_0}^{min(s_1,s_2)} \Phi(t_1, s_1)\Phi^\dagger(t_1, s_2)dt_1 + \Phi(t_0, s_1)\mathcal{K}_v(t_0, t_0)\Phi^\dagger(t_0, s_2)\ , \tag{9}$$

where $\mathcal{K}_v(t_0, t_0)$ is the covariance of the initial velocity, $v(t_0)$. The inertial position process is then Gaussian with covariance $\mathcal{K}_p(s_1, s_2) = \int_{t_0}^{s_1}\int_{t_0}^{s_2} \Psi(\tau_1)\mathcal{K}_v(\tau_1, \tau_2)\Psi^\dagger(\tau_2)d\tau_1 d\tau_2$. The covariance function is parameterized by the sequence of airplane orientations thereby demonstrating the fundamental link between tracking unresolved targets and high-resolution recognition algorithms.

The Gibbs' potential associated with that prior density on the set of inertial positions can be written as

$$P_t(\vec{p}_t(M)) = -\frac{1}{2\sigma_1^2}\vec{p}_t(M)^\dagger K_p^{-1}\vec{p}_t(M), \tag{10}$$

where K_p (obtained using covariance function $\mathcal{K}_p$) is the $3n_M \times 3n_M$ covariance matrix of the position vector $\vec{p}_t(M)$, and $n_M = \sum_{m=1}^{M} t^{(m)}$, the total number of track-segments in $\vec{x}_t(M)$.

3.1.2 Prior on Rotational Motion

We have utilized a von-Mises prior on the orientation angles $\vec{\phi} = [\phi_1, \phi_2, \phi_3] \in \mathcal{M}(3)$ (pitch, roll and yaw). For simplification, it is assumed that the three angles are statistically independent of each other (even though they are known to be related). For circular parameters the von-Mises density is analogous to the normal distribution on the real line [22, 23]. The Gibbs' potential associated with the prior on the rotational motion of M targets is given by

$$P_t(\vec{\phi}_t(M)) = \kappa \left(\sum_{m=1}^{M} \sum_{(k \in \mathcal{T}^{(m)})} \sum_{i=1}^{3} cos(\phi_i^{(m)}(k) - \phi_i^{(m)}(k-1)) \right) \tag{11}$$

where $\kappa > 0$ is called the concentration parameter. The prior potential on the complete parameter space $\mathcal{X}_t$ given M becomes

$$P_t(\vec{x}_t(M)) = P_t(\vec{p}_t(M)) + P_t(\vec{\phi}_t(M)) \, .$$

3.2 Data Likelihood

In this section, we derive the likelihood of collected data conditioned on a given set of parameters. There are two sensor types in our problem, a *tracking sensor*, consisting of an array of passive sensors and a range radar, and a *high-resolution imaging* sensor.

3.2.1 Tracking

For azimuth-elevation coordinate tracking, a cross array of isotropic sensors (Figure 4) is assumed as in [13, 14, 15, 16] using the standard narrowband signal model developed in [12]. Accordingly the signal arriving at the sensors is assumed to be in a relatively small band in the frequency spectrum, such that the signal amplitudes remain approximately constant as wavefronts traverse the array with only difference among the signals reaching different elements being due to the relative phase lags. Depending on the geometry of the sensor arrangement, determining the array-manifold, these phase lags are known functions of the source locations. The amplitudes of the arriving signals $\vec{s}(k)$ can be modeled either as unknown deterministic values or as random variables. In this report, we assume the deterministic signal model in which the measurements $\vec{y}_1(k)$ are Gaussian distributed with mean given by the signal component. Accordingly, the expression for sensor response to M signal sources at locations $\vec{p}^{(1)}(k), .., \vec{p}^{(M)}(k)$ with amplitudes

$\vec{s}(k) = [s^1(k)\ ,..,s^M(k)]$ is

$$\vec{y}_1(k) = \sum_{m=1}^{M} d(\vec{p}^{(m)}(k)) 1_{\mathcal{T}(m)}(k) s^{(m)}(k) + \vec{n}_1(k)\ , \tag{12}$$

where $\vec{n}_1(k)$ is a 0-mean complex Gaussian noise vector of the Goodman class with covariance $\sigma_1^2 I$, $1_{\mathcal{T}(m)}(k)$ selects the targets that contribute in the signal at time k, and $\vec{d}(\vec{p}^{(m)}(k))$ is the vandermonde direction vector corresponding to m^{th} target. In a general problem, the signal amplitudes are also to be estimated from the collected data. But we focus on the estimation of the target positions by assuming the signal amplitudes to be known. The ambient noise surrounding the sensor elements is modeled as a white Gaussian process, i.e. the noise samples added to the signal at different sensors or different times are uncorrelated.

The set of tracking data collected up to time t is given by $I_t^{\mathcal{P}}(1) \equiv \{\vec{y}_1(k) : k \in \{1,..,t\}\}$, and the likelihood of these data has the Gibbs' potential

$$L_t^1(\vec{x}_t(M)) \ = \ -\frac{1}{\sigma_1^2} \sum_{k=1}^{t} |\vec{y}_1(k) - \sum_{m=1}^{M} d(\vec{p}^{(m)}(k)) 1_{\mathcal{T}(m)}(k) s^{(m)}(k)|^2\ .$$

3.2.2 Imaging

While the statistical models for high-resolution radar imaging are being incorporated in this problem by others ([24, 25, 26, 27, 28, 29]), all of the results shown here are based on an optical imaging system. In this system, the data are a sequence of 2-D images resulting from projecting the target onto the focal plane of the imaging sensor; i.e., the deformation of the imaging process of ideal targets is assumed here to be far field orthographic projection. This projection $\mathcal{P}(\cdot)$ defines the deterministic operation of imaging the scene containing multiple objects,

$$\mathcal{P} : \Re^{\mathcal{V}} \rightarrow \Re^{\mathcal{L}}\ ,$$

where $\mathcal{V}$ is the imaged space and $\mathcal{L}$ is the discrete lattice on which the 3D volume is projected. This projection is described as follows: we define a scene, or a configuration of multiple generators $g(1), g(2), ..g(M)$ to be the generators placed and oriented according to their associated parameter vector. Then the volume $\mathcal{V}$ containing the generators is projected onto $\mathcal{L}$ using far field orthographic assumptions as shown in the Figure 5. Since the parameter set $\vec{x}_t(M)$ completely determines the imaged volume, we can also write the projection as an operation from the parameter space to $\Re^{\mathcal{L}}$, i.e. $\mathcal{P} : \mathcal{X}_t \rightarrow \Re^{\mathcal{L}}$.

We choose the lattice to be an $\mathcal{L} = 64 \times 64$ array implying the imaging data at time k form the set of 64×64 grey scale pixel values $\vec{y}_2(k) \in [0, 255]^{64 \times 64}$. For simulations, $\mathcal{P}$ is implemented using the Silicon Graphics imaging system. For the implementation presented here, a Gaussian

noise model was used, with the measured data for the set of M targets having mean given by the projection of M targets. For the scene containing M targets the likelihood potential becomes

$$L_t^2(\vec{x}_t(M)) = -\frac{1}{2\sigma_2^2}\sum_{k=1}^{t}||\vec{y}_2(k) - \mathcal{P}(\vec{x}_t(M))||^2 \tag{13}$$

where $||\cdot||$ represents matrix 2-norm, and σ_2^2 is the noise power.

The imaging data set up to time t are given by $I_t^{\mathcal{D}}(2) \equiv \{\vec{y}_2(k) : k \in \{1,.,t\}\}$, and the complete data set becomes $I_t^{\mathcal{D}} = \{I_t^{\mathcal{D}}(1), I_t^{\mathcal{D}}(2)\}$. The combined data likelihood has potential

$$\begin{aligned} L_t(\vec{x}_t(M)) &= L_t^1(\vec{x}_t(M)) + L_t^2(\vec{x}_t(M)) \\ &= -\frac{1}{\sigma_1^2}\sum_{k=1}^{t}|\vec{y}_1(k) - \sum_{m=1}^{M} d(\vec{p}^{(m)}(k))1_{\mathcal{T}^{(m)}}(k)s^{(m)}(k)|^2 \\ &- \frac{1}{2\sigma_2^2}\sum_{k=1}^{t}\{||\vec{y}_2(k) - \mathcal{P}(\vec{x}_t(M))||^2\} \ . \end{aligned}$$

Shown in Figure 6 are four samples of the imaging data for a single target scene. It shows the ideal projected onto a 2-D lattice with additive noise at four different time instants. Figure 7 shows the spatial power spectrum of the tracking data generated using the *minimum variance distortionless response* (MVDR) [30] spectral analysis at four instants of time, plotted in the azimuth-elevation plane (bright is low power, dark is high power).

Remark: For observing the range locations of the targets, a range radar is assumed with the observations modeled as normally distributed with mean $|\vec{p}(k)|$, the 2-norm of the position vector at time k.

3.3 Bayesian Posterior

The posterior distribution is obtained as the product of the data likelihood and the prior distribution using Bayes' rule. The posterior distribution in Gibbs' form becomes

$$\begin{aligned} \pi_t(\vec{x}_t(M)|M) &= \frac{1}{\mathcal{Z}(M)}e^{-E_t(\vec{x}_t(M)|M)} \\ &= \frac{1}{\mathcal{Z}(M)}e^{-(P_t(\vec{x}_t(M))+L_t(\vec{x}_t(M)))} \ , \end{aligned}$$

where $L_t(\vec{x}_t(M))$ is the potential associated with data likelihood, and $P_t(\vec{x}_t(M))$ is the potential associated with the prior distribution on the parameter space $\mathcal{X}_t(M)$.

So far we have defined a family of posterior distributions $\mu_t(\cdot|M)$ each associated with a

subspace $\mathcal{X}_t(M)$ such that

$$\mu_t(d\vec{x}_t(M)|M) = \frac{1}{\mathcal{Z}(M)} e^{-E_t(\vec{x}_t(M)|M)} d\vec{x}_M$$

where $d\vec{x}_M$ is the appropriate Lebesgue measure associated with the space in which $\vec{x}_t(M)$ is an element. For arbitrary $x_t \in \mathcal{X}_t$, define the potential $E_t(x|M) = 0$, for $x_t \notin \mathcal{X}_t(M)$. Then the posterior measure $\mu_t(\cdot)$ with density $\pi_t(\cdot)$ is in Gibb's form according to

$$\mu_t(dx) = \frac{\sum_{M=0}^{\infty} e^{-E_t(x|M)} 1_{\mathcal{X}_t(M)}(x)}{\mathcal{Z}_t} dx \ , \tag{14}$$

with the normalizer $\mathcal{Z}_t = \sum_{M=0}^{\infty} \int_{\mathcal{X}_t(M)} e^{-E_t(x|M)} dx$.

Having derived a posterior measure over the complete parameter space we now describe an estimation procedure based on the jump-diffusion random sampling algorithm.

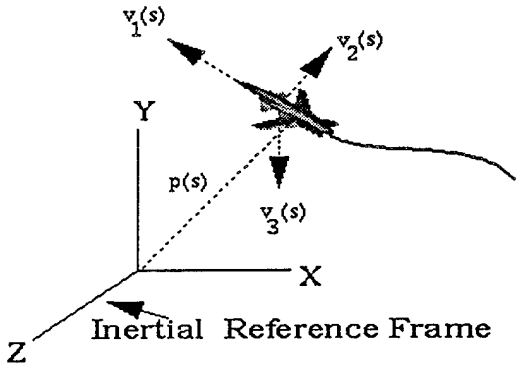


Figure 3: The observed target located at position $\vec{p}(s)$, oriented at $\vec{\phi}(s)$ with body frame velocities $\vec{v}(s)$.

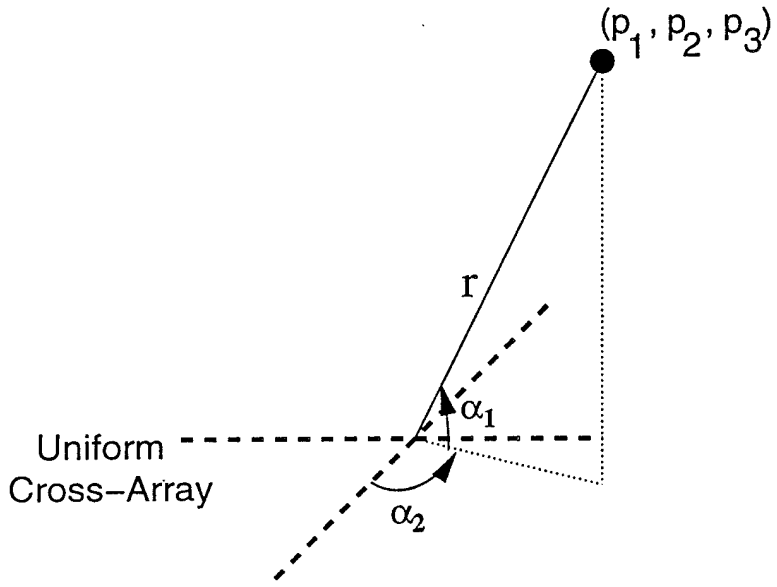


Figure 4: The figure displays the cross array of isotropic sensors at half wavelength spacing, which observes the angular location of the target.

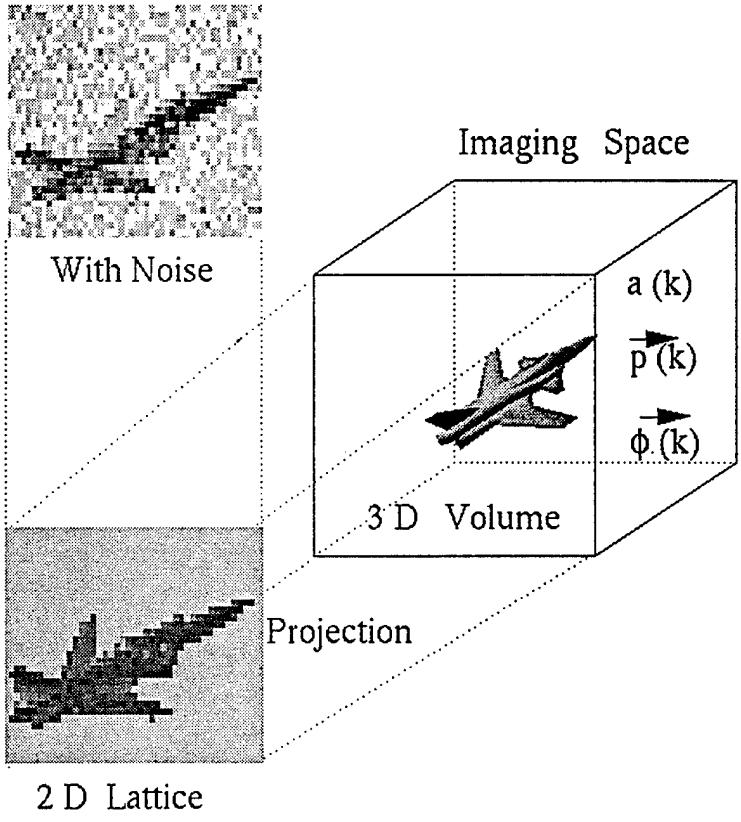


Figure 5: The projection transformation $\mathcal{P}$ converting a 3D volume into a 2D image on a discrete lattice $\mathcal{L}$ for a single target.

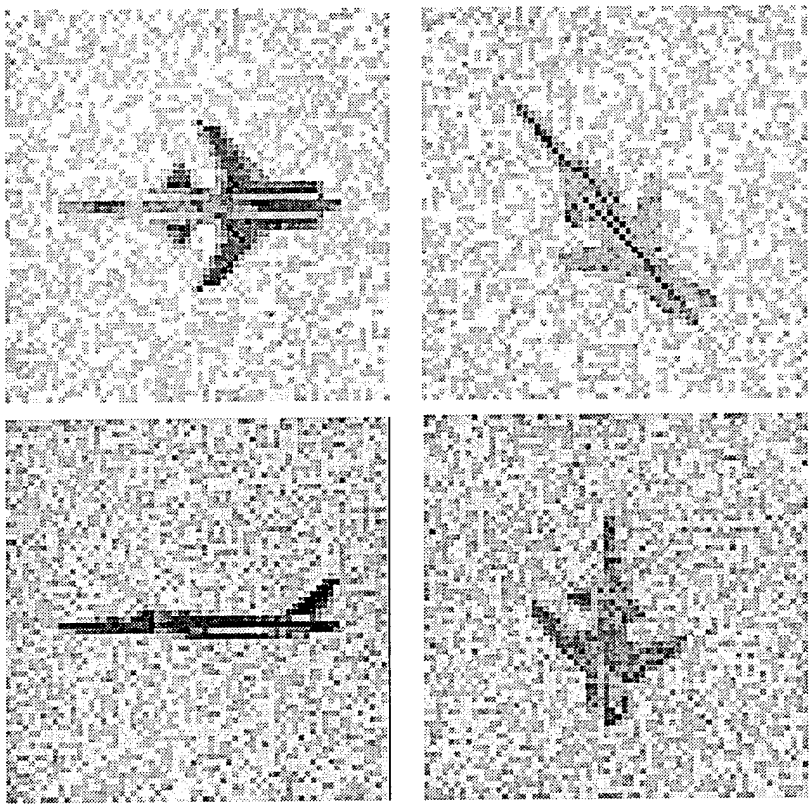


Figure 6: The figure shows the target projected onto a 64×64 2-D lattice with additive noise at four different time instants.

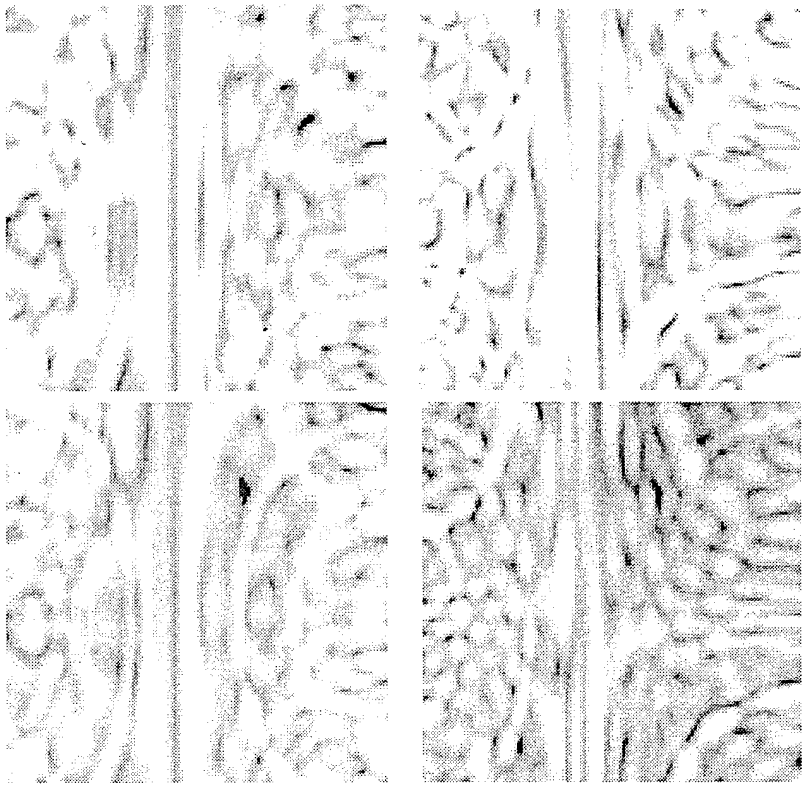


Figure 7: The figure shows the azimuth-elevation spatial power spectrum of the narrowband tracking data generated via MVDR method at four different instants of time (bright is low power, dark is high power).

4 Estimation Through Random Sampling

4.1 Random Sampling

Random sampling from a probability measure over a state space refers to drawing the elements from that space according to that fixed probability measure. The samples are generated via a Markov process which visits the elements of the state space with the frequencies proportional to that probability measure. This implies that the empirical averages generated from the samples of the Markov process converge to their conditional means under the given density.

4.1.1 Why Random Sampling

The estimates are obtained on the basis of *MMSE* criterion which minimizes the cost function $\mathcal{E}\{(\vec{x}_t(M) - \hat{\vec{x}}_t(M))^2 | I_t^{\mathcal{D}}\}$. So the problem is to search for $\hat{\vec{x}}_t(M)_{MS}$ such that

$$\hat{\vec{x}}_t(M)_{MS} = arg \min_{\hat{\vec{x}}_t(M) \in \mathcal{X}_t} \mathcal{E}\{(\vec{x}_t(M) - \hat{\vec{x}}_t(M))^2 | I_t^{\mathcal{D}}\} ,$$

where $\mathcal{E}$ stands for the expectation operator. It is well known that this minimizer is given by the conditional mean ([31]) under the posterior density, i.e.

$$\hat{\vec{x}}_t(M)_{MS} = \mathcal{E}\{\vec{x}_t(M) | I_t^{\mathcal{D}}\}$$

In most problems of practical relevance it is difficult to analytically solve for this conditional mean because of the complicated posterior densities involved. Therefore we define a random sampling mechanism which draws samples from the posterior density such that their averages converge to the conditional mean. This sampling mechanism is based on the jump-diffusion processes.

4.2 Jump-Diffusion Sampling Algorithm

The sampling process is essentially a search for the features which best conform to the given data set. These features can be of discrete nature, e.g. the number of targets, and target type or they can live in continuous spaces, e.g. the target positions and orientations. Accordingly, there are two components in the discovery process which account for these two kinds of feature variabilities. The jump process involves discrete moves over non-connected subspaces searching for the discrete features while the diffusion component performs continuous stochastic gradients estimating the continuous features.

Our approach is to construct a jump-diffusion Markov process, following the analysis outlined in [4], having the limiting property that it converges in distribution to the Bayes posterior. This implies that the time samples of the Markov process visit the configurations with high proba-

bility more often. Following the *jump-diffusion* dynamics the process (i) on random exponential times jumps from one of the countably infinite set of subspaces to another estimating discrete parameters, and (ii) between jumps it performs diffusion following the S.D.E.'s appropriate for subspace it is in.

As mentioned previously, the posterior density changes at each observation time due to the addition of one more data sample to the data set. Therefore, at any given time t, the sampling process generates samples from the posterior density having the Gibbs energy $E_t(\vec{x}_t(M))$. The jump-diffusion Markov process $\{X(s), s \geq 0\}$ samples from the posterior density $\pi_t(\vec{x}_t(M))$ defined over the full parameter space $\mathcal{X}_t$ as follows.

To simplify the following analysis we introduce some additional notation. Define a deletion operator $\wp$ for deleting elements from the present configuration such that $\wp_T^{(j)}$ deletes the j^{th} track while $\wp_S^{(j)}$ removes the last track segment of the j^{th} track, i.e.,

$$\begin{aligned}
\wp_T^{(j)} \;&:\; \left(\mathcal{M}(3) \times \Re^3 \times \mathcal{A}\right)^{\left(\sum_{m=1}^{M} t^{(m)}\right)} \times \aleph^M \rightarrow \\
&\qquad\qquad \left(\mathcal{M}(3) \times \Re^3 \times \mathcal{A}\right)^{\left(\sum_{m\neq j}^{M} t^{(m)}\right)} \times \aleph^{M-1}\,, \\
\wp_S^{(j)} \;&:\; \left(\mathcal{M}(3) \times \Re^3 \times \mathcal{A}\right)^{\left(\sum_{m=1}^{M} t^{(m)}\right)} \times \aleph^M \rightarrow \\
&\qquad\qquad \left(\mathcal{M}(3) \times \Re^3 \times \mathcal{A}\right)^{\left(\sum_{m=1}^{M} t^{(m)}\right)-1} \times \aleph^M\,.
\end{aligned}$$

Also, $\oplus_j$ stands for the addition of elements in the present configuration at j^{th} track, i.e. $\vec{x}_t(M) \oplus_j \vec{y}_t(1)$ represents an $M+1$ track configuration formed by adding $\vec{y}_t(1)$ to $\vec{x}_t(M)$ at the j^{th} location. Similarly $\vec{x}_t(M) \oplus_j y^{(j)}$ signifies addition of a segment to the j^{th} track of $\vec{x}_t(M)$.

4.2.1 Jump Process

The jump process deals with the assignment of tracks and the choice of model order in two ways. First, the individual tracks are developed by probabilistically placing the track-segments sequentially in the associated track configuration. Secondly, the jump process moves among the subspaces of variable numbers of tracks via the addition and deletion of tracks.

For hypothesizing the existence of new tracks and the disappearance of faulty track hypotheses as well as growing and shrinking tracks, we use classical ideas associated with birth-death processes, where a birth corresponds to newly hypothesized track/track-segment in the scene and a death the removal of one. These births/deaths are two of a family of *simple jump moves*, others corresponding to splitting and fusion of tracks, with the simple moves transforming one model to another. The jump transformations are applied discontinuously and drawn probabilis-

tically from a rich family of transformations. The jumps take one model into another satisfying the condition that given any two models of the dimensions M, M', it should be possible to find a finite chain of transitions leading from one to the other.

We allow only those jump moves which result in the following types of transformations through parameter space

$$\begin{aligned}
\text{Addition of track} &: \vec{x}_t(M) \rightarrow \vec{x}_t(M) \oplus_j \vec{y}_t(1), \\
\text{Addition of track-segment} &: \vec{x}_t(M) \rightarrow \vec{x}_t(M) \oplus_j y^{(j)} , \\
\text{Deletion of track} &: \vec{x}_t(M) \rightarrow \wp_T^{(j)} \vec{x}_t(M) , \\
\text{Deletion of track-segment} &: \vec{x}_t(M) \rightarrow \wp_S^{(j)} \vec{x}_t(M) .
\end{aligned}$$

It should be noted that the addition of only unit length tracks is allowed. Let $\mathcal{T}^1(\vec{x}_t(M))$ be the set of configuration types that can be reached from $\vec{x}_t(M)$ in one jump move, i.e.

$$\mathcal{T}^1(\vec{x}_t(M)) = \{\vec{x}_t(M) \oplus_j \vec{y}_t(1), \vec{x}_t(M) \oplus_j y^{(j)}, \wp_T^{(j)} \vec{x}_t(M), \wp_S^{(j)} \vec{x}_t(M)\},$$

$\mathcal{X}_t(\mathcal{T}^1(\vec{x}_t(M)))$ being the space containing the configurations of these types. The discrete jump moves are performed on the basis of following jump parameters defined for $a, b \in \mathcal{X}_t$ as:

- $q(a, db)$: the transition measure from the configuration a to an infinitesimal neighborhood of b.
- $q(a)$: the intensity of jumping out of configuration a, $q(a) = \int_{\mathcal{X}_t(\mathcal{T}^1(a))} q(a, db)$.
- $Q(a, db)$: the transition probability from a to db. $Q(a, db) = \frac{q(a,db)}{\int_{\mathcal{X}_t(\mathcal{T}^1(a)} q(a,db))}$.

The transition measures for the feasible jump moves are given by

$$\begin{aligned}
q(\vec{x}_t(M), d\vec{y}_t(M+1)) &= \sum_{j=1}^{M+1} q_T^b(\vec{x}_t(M), \vec{y}(M+1)) \delta_{\vec{x}_t(M)} \Big(d(\wp_T^{(j)} \vec{y}(M+1)) \Big) d\vec{y}(1) , \\
q(\vec{x}_t(M), d\vec{y}_t(M)) &= \sum_{j=1}^{M} q_S^b(\vec{x}_t(M), \vec{y}_t(M)) \delta_{\vec{x}_t(M)} \Big(d\Big(\wp_S^{(j)} \vec{y}_t(M) \Big) \Big) dy \\
&+ \sum_{j=1}^{M} q_S^d(\vec{x}_t(M), \vec{y}_t(M)) \delta_{\wp_S^{(j)} \vec{x}_t(M)} (d\vec{y}_t(M)) , \\
q(\vec{x}_t(M), d\vec{y}_t(M-1)) &= \sum_{j=1}^{M} q_T^d(\vec{x}_t(M), \vec{y}_t(M-1)) \delta_{\wp_T^{(j)} \vec{x}_t(M)} (d\vec{y}_t(M-1)) \qquad (15)
\end{aligned}$$

and the intensity of jumping out of $\vec{x}_t(M)$ is given by

$$\begin{aligned} q(\vec{x}_t(M)) &= \sum_{j=1}^{M+1} \int_{\mathcal{X}_t(1)} q_T^b\Big(\vec{x}_t(M), \vec{x}_t(M) \oplus_j \vec{y}_t(1)\Big) d(\vec{y}_t(1)) \\ &+ \sum_{j=1}^{M} q_T^d\Big(\vec{x}_t(M), \wp_T^{(j)} \vec{x}_t(M)\Big) \\ &+ \sum_{j=1}^{M} \int_{R^3 \times \mathcal{M}(3) \times \mathcal{A}} q_S^b\Big(\vec{x}_t(M), \vec{x}_t(M) \oplus y^{(j)}\Big) dy \\ &+ \sum_{j=1}^{M} q_S^d\Big(\vec{x}_t(M), \wp_S^{(j)} \vec{x}_t(M)\Big). \end{aligned} \tag{16}$$

The birth/death intensities $q_T^b, q_T^d, q_S^b, q_S^d$ can be derived from the posterior measures of the present and the candidate configurations in two ways. One is analogous to a Gibbs sampling type algorithm while the other is analogous to a Metropolis type acceptance/rejection algorithm. The intensities obtained from the first method are given by,

- Gibb's sampling:

$$\begin{aligned} q_T^b(\vec{x}_t(M), \vec{x}_t(M) \oplus_j \vec{y}_t(1)) &= \pi(\vec{x}_t(M) \oplus_j \vec{y}_t(1)) \\ q_S^b(\vec{x}_t(M), \vec{x}_t(M) \oplus_j y^{(j)}) &= \pi(\vec{x}_t(M) \oplus_j y^{(j)})) \\ q_T^d(\vec{x}_t(M), \wp_T^{(j)}(\vec{x}_t(M))) &= \pi(\wp_T^{(j)}(\vec{x}_t(M))) \\ q_S^d(\vec{x}_t(M), \wp_S^{(j)}(\vec{x}_t(M))) &= \pi(\wp_S^{(j)}(\vec{x}_t(M))) \end{aligned}$$

These expressions provide the birth/death intensities for constructing the jump process having jump moves derived from Gibb's sampling. Our implementation is based on the jump moves of second type which represents a modification of the Metropolis algorithm introduced in 1953 [32]. This algorithm is implemented through the following steps.

Metropolis Based Jump-Algorithm

1. Generate independent exponential r.v.'s $u_1, u_2, ..$ with the intensity λ, where λ is the average number of diffusion cycles for every jump move.

2. At time $t_i = \sum_{j=1}^{i} u_j$ draw a candidate $\vec{y}_t(M')$ from the prior (e.g by using the equations of motion).

3. Compute $L_t(\vec{y}_t(M'))$.
   If $[L_t(\vec{x}_t(M)) - L_t(\vec{y}_t(M'))] > 0$,
   go to $\vec{y}_t(M')$.
   else
   go to $\vec{y}_t(M')$ with the probability $e^{-[L_t(\vec{y}_t(M'))-L_t(\vec{x}_t(M))]}$.

4. Repeat step 2.

The corresponding birth/death intensities for $\vec{y}_t(M') \in \mathcal{X}(\mathcal{T}^1(\vec{x}_t(M)))$ are given by

$$
\begin{aligned}
q_T^b(\vec{x}_t(M), \vec{x}_t(M) \oplus_j \vec{y}_t(1)) &\doteq \frac{1}{4(M+1)} e^{-[L(\vec{x}_t(M)\oplus_j \vec{y}_t(1))-L(\vec{x}_t(M))]_+} \frac{e^{-P(\vec{y}_t(1))}}{Z_T(1)}, \\
q_S^b(\vec{x}_t(M), \vec{x}_t(M) \oplus_j y^{(j)}) &= \frac{1}{4M} e^{-[L(\vec{x}_t(M)\oplus_j y^{(j)})-L(\vec{x}_t(M))]_+} \frac{e^{-P(y^{(j)})}}{Z_S(1)}, \\
q_T^d(\vec{x}_t(M), \wp_S^{(j)} \vec{x}_t(M)) &= \frac{1}{4M} \frac{e^{-[L(\wp_S^{(j)}\vec{x}_t(M))-L(\vec{x}_t(M))]_+}}{Z_T(1)}, \\
q_S^d(\vec{x}_t(M), \wp_S^{(j)} \vec{x}_t(M)) &= \frac{1}{4M} \frac{e^{-[L(\wp_S^{(j)}\vec{x}_t(M))-L(\vec{x}_t(M))]_+}}{Z_S(1)}, \qquad (17)
\end{aligned}
$$

where $Z_T(1), Z_S(1)$ are the partition functions for the prior densities on single track and single track-segment configurations respectively. In between jump moves the process stays in the current subspace and performs stochastic gradient search via diffusion process.

4.2.2 Diffusion Process

The diffusion process contributes in the search of the features which lie in the continuous space. It is a sample path continuous process which performs a randomized gradient search on the posterior potential $E_t(\vec{x}_t(M))$ in the current subspace $\mathcal{X}_t(M)$ according to Langevin's stochastic differential equation (SDE).

A diffusion is completely defined by its infinitesimal mean and variance. In this approach we generate a diffusion by assigning the gradients of the posterior energy to be the infinitesimal mean. For the sub-space containing M tracks the diffusion flows through the manifold $\mathcal{X}_t(M)$ estimating the continuous parameters: the target's positions $\vec{p} \in \Re^3$ and orientations $\vec{\phi} \in \mathcal{M}(3)$. Define $n_M = \sum_{m=1}^{M} t^{(m)}$, the total track segments in $\vec{x}_t(M)$ and associate with the first

$3n_M$ components of the S.D.E. $X(s)$ the flow through $\mathcal{M}(3)^{n_M}$ to estimate the orientations as described in [33], and the last $3n_M$ components the flow through $\Re^{(3n_M)}$ to estimate the positions. Then the diffusion $X(s)$ satisfies the following vector S.D.E.:

$$X_1(s) = \left[X_1(0) + \int_0^s -\frac{1}{2}\nabla_1 E_t(X(\tau))d\tau + W_1(s)\right]_{\text{mod } 2\pi}, \quad (18)$$

$$X_2(s) = X_2(0) + \int_0^s -\frac{1}{2}\nabla_2 E_t(X(\tau))d\tau + W_2(s), \quad (19)$$

where $[\cdot]_{\text{mod } 2\pi}$ is taken componentwise, $(X(s) = [X_1(s), X_2(s)])$, and $W_1(s), W_2(s)$ are the standard vector Wiener processes of dimensions $3n_M$, and ∇_1, ∇_2 are the gradients with respect to the vectors $\vec{p}_t(M), \vec{\phi}_t(M)$.

4.2.3 Ergodic Result

Now we present an important result on the ergodic properties of jump-diffusion process. This result verifies the claim that the jump-diffusion process constructed above samples from the posterior distribution $\pi_t(\vec{x}_t(M))$.

Theorem 1 *If the jump diffusion process $X(s)$ has the properties that:*
(a) the diffusion $X(s)$ within any subspace satisfies the S.D.E. of Eqns (18,19)
(b) the birth/death intensities defined by Eqn (17) generate the jump process with the parameters given by Eqns (15,16)

then $X(n\Delta)$ converges in variation norm to $\mu_t(d\vec{x}_t(M)) = \pi_t(\vec{x}_t(M))d\vec{x}_t(M)$.

Proof: See Appendix 1.

In Eqn (19), the term $\nabla_2 E_t(X(s))$ has two gradient components: the likelihood gradient $\nabla_2 L_t(X(s))$ and the prior gradient $\nabla_2 P(X(s))$. Since we use a Gaussian prior on target positions the gradient of prior potential is given by $K_p^{-1}\vec{p}_t(M)$. Clearly, for the dynamic scenarios with changing configurations and therefore changing covariances the matrix-inverse computation gets intensive. We now present an alternate diffusion process which also samples from the same density and is computationally far less intensive. This diffusion follows the SDE,

$$X_2(s) = X_2(0) + \int_0^s -\frac{1}{2}\left(K_p\nabla_2 L_t(X(\tau)) + X(\tau)\right)d\tau + \sqrt{K_p}dW_2(s) \quad (20)$$

where K_p is the $3n_M \times 3n_M$ covariance matrix of the position vector $\vec{p}_t(M)$ and $W_2(s)$ is the standard Wiener process of dimension $3n_M$. The following theorem concludes that this diffusion also samples from the same posterior distribution.

Corollary 1 *The modified jump-diffusion process with the properties that:*
(a) the diffusion X(s) within any subspace satisfies the S.D.E. of Eqns (18,20),

(b) the birth/death intensities defined by Eqn (17) generate the jump process with the parameters given by Eqns (15,16),

then $X(n\Delta)$ *converges in variation norm to* $\mu_t(d\vec{x}_t(M)) = \pi_t(\vec{x}_t(M))d\vec{x}_t(M)$.

Proof: See Appendix 2.

5 Implementation and Results

Now we present the implementation of a jump-diffusion algorithm for estimating the motion of a single target, i.e. M = 1. The algorithm was jointly implemented using a Silicon Graphics workstation for data generation and visualization, and a massively parallel 4096 processor SIMD DECmpp machine for implementing the track-recognition algorithm.

The parameter set describing the target configuration for the observation interval $\{1, ..t\}$, is

$$\begin{aligned} \vec{x}_t(1) &= \{\vec{x}^{(1)}(k) : k \in \mathcal{T}^{(1)}\} \\ &\in \mathcal{X}_t(1) \equiv \bigcup_{t^{(1)}=1}^{t} \left(\mathcal{M}(3) \times \Re^3 \times \mathcal{A}\right)^{t^{(1)}} \times \aleph \end{aligned}$$

For estimating the object orientation, the orientation space $\mathcal{M}(3)$ was uniformly sampled and 64×64 pixel 2D projections $\mathcal{P}(\cdot)$ of the 3D surface of the target at sampled orientations were generated and stored. This set of templates form the object space over which the recognition is performed, i.e the estimates are selected from this set.

5.1 Data Simulation

The *flight simulator* software on Silicon Graphics workstation was utilized to generate parameterized airplane paths. These path coordinates were then used in data simulation modules on mpp and Silicon Graphics to obtain data sets corresponding to the narrowband tracking array and the optical imaging sensor, respectively. The tracking data consists of a 64 length complex vector sampled at the cross-array $\vec{y}_1(k)$ and a real number corresponding to the range data for the target $r(k)$ at each time index k. The array geometry corresponds to a 64-element cross-array of isotropic sensors located at half-wavelength spacing. The tracking data $\{y_1(k)\}$ is a 64-element complex vector with mean $d(p(k))s(k)$ and additive complex Gaussian white noise of the Goodman's class. The direction vector corresponds to the array takes the form

$$[e^{-\frac{i31}{2}\lambda_1(k)},\ e^{-\frac{i29}{2}\lambda_1(k)}\ ,...,e^{\frac{i31}{2}\lambda_1(k)},\ e^{-\frac{i31}{2}\lambda_2(k)},e^{-\frac{i29}{2}\lambda_2(k)}\ ,..,\ e^{\frac{i31}{2}\lambda_2(k)}]^T\ ,\quad i=\sqrt{-1}\ ,$$

where $\lambda_1(k) = \pi cos(\alpha_1(k))sin(\alpha_2(k))$, and $\lambda_2(k) = \pi cos(\alpha_1(k))sin(\alpha_2(k))$, $\alpha_1(k), \alpha_2(k)$ is the azimuth, elevation angles of the target position $p(k)$.

The imaging data, generated by optical imaging of the space around the estimated position of the target using Silicon Graphics, consists of a 64×64 matrix $\vec{y}_2(k)$ of grey scale pixel values for each observation. The data set is then transferred to the mpp where the orientation and target type estimation is performed. In the results presented we have used high signal to noise ratio for both the tracking and imaging data sets. The noise in the data model was generated using the Gaussian random number generator on the machines.

5.2 Estimation Algorithm

Now we describe the jump-diffusion algorithm for estimating the single track scene. The estimation algorithm proceeds by births and deaths of track segments at random times through discrete jump moves. At any given time t the jump-diffusion algorithm is run to generate samples from the posterior distribution given by $\pi_t(\cdot)$. This simulation is performed till the next data set arrives when the algorithm starts sampling from $\pi_{t+1}(\cdot)$ and so on. The two possible jump moves involve the following transformations in the parameter space,

$$\begin{aligned}\vec{x}_t(1) &\rightarrow \vec{x}_t(1) \oplus_1 y^{(1)} \in \left(\Re^3 \times \mathcal{M}(3) \times \mathcal{A}\right)^{t^{(1)}+1} \times \aleph \,, \\ \vec{x}_t(1) &\rightarrow \wp_S^{(1)} \vec{x}_t(1) \in \left(\Re^3 \times \mathcal{M}(3) \times \mathcal{A}\right)^{t^{(1)}-1} \times \aleph \,.\end{aligned}$$

where $\vec{x}_t(1) \in (\Re^3 \times \mathcal{M}(3) \times \mathcal{A})^{t^{(1)}} \times \aleph$. We assign equal probabilities for selecting one of the two options with the actual moves being performed on the basis of posterior energies. Notice that for the option to delete the track segment there is only one candidate configuration but for adding the track segment the candidates are numerous corresponding to all possible values of $y^{(1)}$. In that case, using Metropolis based jump moves the algorithm candidates are generated and selected as follows. The vector differential equation (Eqn 7), describing the motion in the body frame coordinates, can be written in discrete form as

$$\vec{v}(k+1) = (A(k) - I_3)\vec{v}(k) + \vec{f}(k) \,.$$

Suppose the track is estimated up to the $(k+1)^{st}$ stage and the $k+2$ position (or equivalently the $k+1$st velocity component) is to be found. The estimated velocity profile for times $1, 2, .., k$ is available. A sample from $N(0, \sigma_0^2 I_3)$ is substituted for the force vector and the difference equation is solved for $\vec{v}(k+1)$ using the estimated rotational motion and the previous velocity estimate. Using the velocity-position transformation (Eqn 4) this $\vec{v}(k+1)$ provides a candidate for the inertial position $\vec{p}(k+2)$. The orientation and target type components for the candidate segment are chosen to be the same as the previous segment. This new track segment is selected and added to the track estimate according to the Metropolis type jump algorithm (discussed in section 4). The likelihood potential of this candidate, $L(\vec{x}_t(1) \oplus_1 y^{(1)})$, is compared to the likelihood potential of the present estimate, $L(\vec{x}_t(1)$. If $L(\vec{x}_t(1) > L(\vec{x}_t(1) \oplus_1 y^{(1)})$, then the segment $y^{(1)}$ is added to the track $\vec{x}_t(1)$ otherwise it is added with the probability $e^{-(L(\vec{x}_t(1) \oplus_1 y^{(1)}) - L(\vec{x}_t(1))}$.

Then the algorithm adjusts the positions and orientations in the estimated track following the gradients of the posterior according to the diffusion equations until the next jump move is performed. The average number of diffusion cycles per jump move is given by the parameter λ, the mean of exponential times separating jump moves. It should be noted that the gradients

of imaging data likelihood involves the derivative of the projection transform which cannot be derived analytically. These are approximated numerically at the sample orientations by taking the difference of adjacent pre-stored templates at that orientations, scaled by the step size of parameter variation. The gradients of the tracking data likelihood are derived analytically using the chain rule for coordinate transformations. The diffusion process is simulated via the discrete equations corresponding to the SDE's 18,20 given by

$$\begin{aligned} X_1((n+1)\epsilon) &= [X_1(n\epsilon) - \frac{1}{2}\nabla_1 E_t(X(n\epsilon)) + [W_1((n+1)\epsilon) - W_1(n\epsilon)]]_{mod\ 2\pi} \\ X_2((n+1)\epsilon) &= X_2(n\epsilon) - \frac{1}{2}\nabla_2 E_t(X(n\epsilon)) + [W_2((n+1)\epsilon) - W_2(n\epsilon)] \end{aligned}$$

Figure 8 shows three successive stages of the algorithm for estimating a portion of the true track

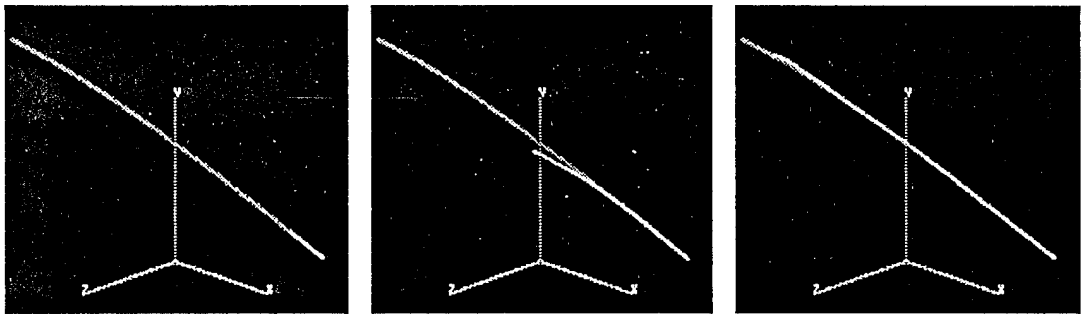


Figure 8: Jump-diffusion estimation of a portion of the track. The true track is drawn in grey while the estimates overlap in white. The estimation proceeds via a sequence of jump moves with the diffusion cycles performed between moves.

shown in grey with the estimates overlapping in white. The algorithm proceeds via sequence of jump moves corresponding to the births of track-segments and adjustment of the track-estimates between the jumps via the diffusion algorithm.

Shown in Fig 9 is the result for the complete track estimation algorithm. The left panel shows the simulation environment for the implementation. The mesh represents the ground supporting the inertial frame of reference and the sensor systems while the grey track represents the parameterized plane path generated from the flight simulator on Silicon Graphics and used as true track in the simulations. In the right panel the estimated track is shown in white overlapping the true track obtained via the jump-diffusion algorithm.

5.2.1 Parallel Processing

The DECmpp has a 64×64 mesh of processors each of which can simultaneously operate on a matrix of up to 64×64 data elements. In this application a large part of the computation involves

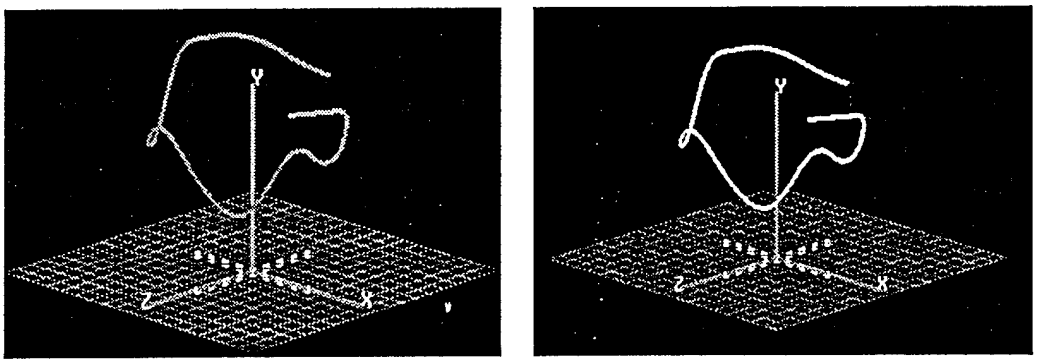


Figure 9: 3D track estimation: The left panel shows the actual track drawn in gray with the mesh representing ground supporting the observation system in the inertial frame of reference. The right panel displays the results from the single track estimation with the estimates drawn in white..

simultaneous operations like coordinate transformations, trigonometric operations, and matrix computations on arrays of data. These operations along with global summation and processor communication are efficiently implemented on a machine like the DECmpp. The choice of sizes of 64-length tracking data vector and 64×64 imaging data allows convenient mapping of the problem onto the 64×64 processor array of mpp.

5.2.2 Remote Visualization

Even though the massively parallel machine is ideal for implementing estimation algorithm it doesn't have adequate graphical resources to provide good display of results. The Silicon Graphics workstation is well suited for 3-D visualization of the actual flight path and the estimated path. Therefore we distribute the tasks across various platforms to make use of advance computing and graphical resources that are not available on any one machine. In fact there could be multiple visualization nodes to address various aspects of the implementation. Also the distributed computation implies more efficient implementation as the tasks are shared by various machines.

The communication between machines demands pipeline of high speed network for data flow. This was implemented using TCP/IP sockets on ethernet to which the mpp and SGI machines were connected. The algorithm performs the computation continuously while feeding the estimates onto the network at regular intervals. The estimates are received by SGI and fed into a visualization program to display the results in a desired way.

Appendix

A Proof of Theorem 1

Proof: This analysis is carried out for a fixed t so we drop the subscript t without any ambiguity. There are essentially two points to the proof: (1) showing that $\pi(\vec{x}(M))$ is an invariant density of the process, and (2) verifying that the process is irreducible and therefore $\pi(\vec{x}(M))$ is the unique invariant density. Part (2) follows directly that in [1, 11] using the properties of the jump process and from the fact that the diffusions are each irreducible over their respective subspaces. In part (1) we need to verify the stationarity for both the jump and diffusion components of the Markov process. The generator, or backward Kolmogoroff operator, for the jump-diffusion process (denote it as $A = A^d + A^j$ (diffusion+jump)) characterizes the stationary density in that $\pi(\vec{x}(M))$ is stationary for the jump-diffusion if and only if $\int Af(\vec{x}(M))\pi(\vec{x}(M))d\vec{x}(M) = 0$ for all f in the domain of A, $\mathcal{D}(A)$.

The diffusion process has two components corresponding to the S.D.E on the multi-dimensional torus (18) and the S.D.E. on the Euclidean space of target positions (19). To prove invariance of $\pi(\vec{x}(M))$ for the diffusion on torus we use results from [11] on invariant distributions of S.D.E.'s on linear manifolds, in particular the multi-dimensional Torus. The stationarity condition is verified for the Euclidean component as follows. Define a set of functions which forms the domain of the generator A as

$$\mathcal{D}(A) = \{f : f = \sum_{m=0}^{M} 1_{\mathcal{X}(m)} f_m, f_m \in \hat{C}^2(\mathcal{X}(m)), M \geq 0\} \ . \tag{21}$$

Then the infinitesimal generator for diffusion A^d becomes

$$A^d f(\vec{x}(M)) = -\frac{1}{2}(\nabla_2 E(\vec{x}(M)) \circ \nabla_2 f(\vec{x}(M))) + \frac{1}{2}\sum_{i=1}^{3n_M} \frac{\partial^2 f(\vec{x}(M))}{\partial^2(\vec{p}(M))_i} \ , \forall f \in \mathcal{D}(A)$$

where $n_M = \sum_{m=1}^{M}$, total number of track-segments in the parameter set $\vec{x}(M)$, $\circ$ stands for the vector dot-product and the gradients $\nabla_2 E(\vec{x}(M)), \nabla_2 f(\vec{x}(M))$ are w.r.t the position vector $\vec{p}(M)$. Substituting this expression in the integral condition we get

$$\begin{aligned}\int A^d f(\vec{x}(M))\pi(\vec{x}(M))d\vec{x}(M) \ &= \\ &- \int \frac{1}{2}(\nabla E(\vec{x}(M)) \circ \nabla f(\vec{x}(M)) \frac{e^{-E(\vec{x}(M))}}{\mathcal{Z}} d\vec{x}(M) \\ &+ \int \frac{1}{2}\left(\sum_{i=1}^{3n_M} \frac{\partial^2 f(\vec{x}(M))}{\partial^2(\vec{p}(M))_i}\right) \frac{e^{-E(\vec{x}(M))}}{\mathcal{Z}} d\vec{x}(M) \ .\end{aligned}$$

Integration by parts of the second term, with the fact that the function f vanishes at the boundary, results in a term which is negative of the first term. Therefore the given posterior $\pi(\vec{x}(M))$ is the stationary density of the diffusion process.

The generator A^j of a jump process is given by the expression,

$$A^j f(\vec{x}(M)) = q(\vec{x}(M)) \int_{\mathcal{X}(T^1(\vec{x}(M)))} Q(\vec{x}(M), dy)(f(y) - f(\vec{x}(M))) \, .$$

When substituted in the stationarity condition, it provides

$$q(\vec{x}(M))\pi(\vec{x}(M))d\vec{x}(M) = \int_{\mathcal{X}(T^{-1}(\vec{x}(M)))} q(\vec{y}, d\vec{x}(M))\pi(\vec{y})d\vec{y} \, ,$$

which is often called as the *detailed balance condition.* Therefore, the jump parameters should satisfy this equation for the density $\pi(\vec{x}(M))$ to be the stationary density of the jump process. We will prove this condition assuming only birth/death of tracks, the treatment for birth/deaths of track segments being similar. Substituting for the transition measures in the *detailed balance condition* and simplifying we obtain,

$$\begin{aligned}
&\pi(\vec{x}(M))d\vec{x}(M) \, [\sum_{j=1}^{M+1} \int_{\mathcal{X}(1)} q_T^b(\vec{x}(M), \vec{x}(M) \oplus_j \vec{y}(1))d\vec{y}(1) + \sum_{j=1}^{M} q_T^d(\vec{x}(M), \wp_T^{(j)}\vec{x}(M)) \,] \\
= \quad & d\vec{x}(M) \, [\sum_{j=1}^{M+1} \int_{\mathcal{X}(1)} q_T^d(\vec{x}(M) \oplus_j \vec{y}(1), \vec{x}(M))\pi(\vec{x}(M) \oplus_j \vec{y}(1))d\vec{y}(1) \\
& \qquad + \sum_{j=1}^{M} q_T^b(\wp_T^{(j)}\vec{x}(M), \vec{x}(M))\pi(\wp_T^{(j)}\vec{x}(M)) \,] \qquad (22)
\end{aligned}$$

Substituting the values for q_T^b, q_T^d from 17, and analyzing only the j^{th} terms from sums on both sides,

R.H.S.

$$\begin{aligned}
& \frac{d\vec{x}(M)}{Z_T(1)} \, [\int_{\mathcal{X}_1} \frac{1}{4(M+1)} e^{-[L(\vec{x}(M)) - L(\vec{x}(M) \oplus_j \vec{y}(1))]_+} \pi(\vec{x}(M) \oplus_j \vec{y}(1))d\vec{y}(1) \\
& \qquad + \frac{1}{4M} e^{-[L(\vec{x}(M)) - L(\wp_T^{(j)}\vec{x}(M))]_+} e^{-P(\vec{x}^{(j)}(1))} \pi(\wp_T^{(j)}\vec{x}(M)) \,] \\
= \quad & \frac{d\vec{x}(M)}{(\mathcal{Z})(Z_T(1))} \, [\frac{1}{4(M+1)} \int_{\Omega_>} e^{-L(\vec{x}(M))} e^{-P(\vec{x}(M) \oplus_j \vec{y}(1))} d\vec{y}(1) \\
& \qquad + \frac{1}{4(M+1)} \int_{\Omega_\leq} e^{-L(\vec{x}(M) \oplus_j \vec{y}(1))} \, e^{-P(\vec{x}(M) \oplus_j \vec{y}(1))} d\vec{y}(1) \\
& \qquad + \frac{1}{4M} e^{-[L(\vec{x}(M)) - L(\wp_T^{(j)}\vec{x}(M))]_+} e^{-P(\vec{x}^{(j)(1)})} e^{-L(\wp_T^{(j)}\vec{x}(M))} e^{-P(\wp_T^{(j)}\vec{x}(M))} \,] \, ,
\end{aligned} \qquad (23)$$

where

$$\Omega_{>} = \mathcal{X}(1) \bigcap \{\vec{y}(1) : L(\vec{x}(M)) > L(\vec{x}(M) \oplus_j \vec{y}(1))\}$$

$$\Omega_{\leq} = \mathcal{X}(1) \bigcap \{\vec{y}(1) : L(\vec{x}(M)) \leq L(\vec{x}(M) \oplus_j \vec{y}(1))\}$$

L.H.S.

$$\begin{aligned}
&\pi(\vec{x}(M)) \frac{d\vec{x}(M)}{Z_T(1)} \Big[\frac{1}{4(M+1)} \int_{\mathcal{X}(1)} e^{-[L(\vec{x}(M)\oplus_j \vec{y}(1)) - L(\vec{x}(M))]_+} e^{-P(\vec{y}(1))} d\vec{y}(1) \\
&\qquad + \frac{1}{4M} e^{-[L(\wp_T^{(j)} \vec{x}(M)) - L(\vec{x}(M))]_+} \Big] \\
= \; &\frac{d\vec{x}(M)}{(\mathcal{Z})(Z_T(1))} \Big[\frac{1}{4(M+1)} \int_{\Omega_{>}} e^{-L(\vec{x}(M))} e^{-P(\vec{x}(M))} e^{-P(\vec{x}^{(j)}(1))} d\vec{y}(1) \\
&\qquad + \frac{1}{4(M+1)} \int_{\Omega_{\leq}} e^{-L(\vec{x}(M)\oplus_j \vec{y}(1))} e^{-P(\vec{x}(M))} e^{-P(\vec{y}(1))} d\vec{y}(1) \\
&\qquad + \frac{1}{4M} e^{-L(\vec{x}(M))} e^{-P(\vec{x}(M))} e^{-[L(\wp_T^{(j)} \vec{x}(M)) - L(\vec{x}(M))]_+} \Big] \qquad (24)
\end{aligned}$$

Comparing the equations 23, 24 and using the independence of priors, for different tracks, the condition is verified.

B Proof of Corollary 1

The Markov process $X(s)$ described here differs from the one in previous theorem in only the diffusion component. Therefore the proof follows similarly for the jump process while the conditions of stationarity are to be verified for the diffusion part. We need to show that the backward kolmogoroff operator A^d for the diffusion process satisfies the equation $\int A^d f(\vec{x}(M)) \pi(\vec{x}(M)) m(d\vec{x}(M)) = 0$.

The diffusion generated by S.D.E. (Eqn 20) has the associated infinitesimal generator A^d is given by,

$$A^d f(\vec{x}(M)) = -[K_p \nabla_2 L(\vec{x}(M)) + \vec{p}(M)] \circ \nabla_2 f(\vec{x}(M)) + \sum_{i,j=1}^{3n_M} [K_p \hat{\circ} \nabla_2^2 f(\vec{x}(M))]_{ij} \; .$$

Substituting this expression in the stationarity condition,

$$\begin{aligned}
\int A f(\vec{x}(M)) \pi(d\vec{x}(M)) d\vec{x}(M) \; &= \; \int \{ -[K_p \nabla_2 L(\vec{x}(M)) + \vec{p}(M)] \circ \nabla_2 f(\vec{x}(M)) \\
&\sum_{i,j=1}^{3n_M} [K_p \hat{\circ} \nabla_2^2 f(\vec{x}(M))]_{ij} \} \frac{1}{\mathcal{Z}} e^{-(L(\vec{x}(M)) + \frac{1}{2} \vec{p}(M)^t K_p^{-1} \vec{p}(M))} d\vec{x}(M) \; .
\end{aligned}$$

Integrating by parts the second term on right side and using boundary conditions we obtain,

$$\frac{1}{\mathcal{Z}} \int \left[\nabla_2 f(\vec{x}(M))^\dagger \circ \{ K_p \nabla_2 L(\vec{x}(M)) + \vec{p}(M) \} \right] e^{-(L(\vec{x}(M)) + \frac{1}{2}\vec{p}(M)^\dagger K_p^{-1} \vec{p}(M))} d\vec{x}(M) \ ,$$

which is negative of the first term in the equation.

Q.E.D

References

[1] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society*, 56(3), 1994.

[2] B. Gidas. Metropolis type monte-carlo simulation algorithms and simulated annealing. In J. Laurie Snell, editor, *Topics in Contemporary Probability and Its Applications*, April 1995.

[3] J. Besag and P. J. Green. Spatial statistics and bayesian computation. *J. Royal Statistical Society B*, 55:25–38, 1993.

[4] U. Grenander and M. I. Miller. Jump-diffusion processes for abduction and recognition of biological shapes. *Monograph of the Electronic Signals and Systems Research Laboratory*, 1991.

[5] Y. Bar-Shalom and E. Tse. Tracking in cluttered environment with probabilistic data association. *Automatica*, 11:451–460, 1975.

[6] Y. Bar-Shalom and T. E. Fortmànn. *Tracking and Data Association.* Academic Press, 1988.

[7] Editor: Y. Bar-Shalom. *Multitarget-Multisensor Tracking.* Artech House, 1990.

[8] C. R. Rao, C. R. Sastry, and B. Zhou. Tracking the direction of arrival of multiple moving targets. *IEEE Transactions on Acoustics,Speech and Signal Processing*, 42(No. 5), May 1994.

[9] C. R. Sastry, E. W. Kamen, and M. Simaan. An efficient algorithm for tracking angles of arrival of moving targets. *IEEE Transactions on Acoustics,Speech and Signal Processing*, ASSP-39(No.1):242–246, 1991.

[10] C. K. Sword, M. Simaan, and E. W. Kamen. Multiple target angle tracking using using sensor array outputs. *IEEE Trans. on Aerospace and Electronic Systems.*, 26(2):367–373, 1990.

[11] Y. Amit and M.I. Miller. Ergodic properties of jump-diffusion processes. *Monograph of Electronic Signals and Systems Research Laboratory, Washington University, St. Louis*, December,1992.

[12] R. Schmidt. *A signal subspace approach to multiple emitter location and spectral estimation.* Ph.D. Dissertation of Stanford University, Palo Alto, CA., Nov. 1981.

[13] M.I. Miller and D. R. Fuhrmann. Maximum likelihood narrow-band direction finding and the em algorithm. *IEEE Acoust. Speech and Signal Processing*, 38, No.9(38, No.9):560–577, 1990.

[14] A. Srivastava, M. I. Miller, and U. Grenander. Jump-diffusion processes for object tracking and direction finding. In *Proc. 29th Annual Allerton Conference on Communication ,Control and Computing*, pages 563–570, Urbana, Il, October 1991. University of Illinois.

[15] A. Srivastava, N. Cutaia, M. I. Miller, J. A. O'Sullivan, and D. L. Snyder. Multi-target narrowband direction finding and tracking using motion dynamics. In *Proc. 30th Annual Allerton Conference on Communication, Control, and Computing*, pages 279–288, Urbana, Il, October 1992. University of Illinois.

[16] M.I. Miller, R. S. Teichman, A. Srivastava, J.A. O'Sullivan, and D. L. Snyder. Jump-diffusion processes for automated tracking-target recognition. In *1993 Conference on Information Sciences and Systems*, Baltimore, Maryland, March 24-26 1993. Johns Hopkins University.

[17] A. Srivastava, R. S. Teichman, and M. I. Miller. Target tracking and recognition using jump-diffusion processes. In *Proc. ARO's Eleventh Army Conference on Applied Mathematics and Computing*, Carnegie Melon University, Pittsburgh, PA, June 1993.

[18] W. Freiberger and U. Grenander. Computer generated image algebras. *Internation Federation Information Processing*, 68:1397–1404, 1969.

[19] U. Grenander. Advances in pattern theory: The 1985 rietz lecture. *The Annals of Statistics*, 17:1–30, 1985.

[20] Bernard Friedland. *Control System Design : An Introduction To State-Space Methods.* McGraw-Hill Book Company, 1986.

[21] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *J. American Statistical Association*, 1991.

[22] K. V. Mardia. *Statistics of Directional Data.* Academic Press, 1972.

[23] G. S. Watson. *Statistics on Spheres.* John Wiley and Sons, 1983.

[24] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller. The use of maximum-likelihood estimation for forming images of diffuse radar-targets. In *Transactions of SPIE in Advanced Architectures and Algorithms*, San Diego, California, 1987.

[25] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller. The use of maximum-likelihood estimation for forming images of diffuse radar-targets from delay-doppler data. *IEEE Transactions on Information Theory*, pages 536–548, 1989.

[26] J.A. O'Sullivan, P. Moulin, and D.L. Snyder. Cramer-rao bounds for constrained spectrum estimation with application to a problem in radar imaging. In *Proceedings 26th Allerton Conference on Communication, Control, and Computing*, Champaigne, Urbana, October 1988 October 1988. Urbana, IL.

[27] M.I. Miller, D.R. Fuhrmann, J.A. O'Sullivan, and D.L. Snyder. Maximum-likelihood methods for toeplitz covariance estimation and radar imaging. In Simon Haykin, editor, *Advances in Spectrum Estimation.* Prentice-Hall, 1990.

[28] P. Moulin, J.A. O'Sullivan, and D.L. Snyder. A method of sieves for multiresolution spectrum estimation and radar imaging. *to appear in IEEE Transactions on Information Theory*, 1992.

[29] J.A. O'Sullivan, K. C. Du, R. S. Teichman, M.I. Miller, D.L. Snyder, and V.C. Vannicola. Radar target recognition using shape models. In *Proc. 30th Annual Allerton Conference on Communication,Control, and Computing*, pages 515–523, Urbana, Il., 1992. University of Illinois.

[30] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57:1408, 1969.

[31] H.L. Van Trees. *Detection, Estimation, and Modulation Theory, vol. I.* John Wiley, N.Y., 1971.

[32] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Physical Chemistry*, 21:1087, 1953.

[33] Y. Amit. A multiflow approximation to diffusions. *Stochastic Processes and their Applications*, 37(2):213–238, 1991.

[34] I.I. Gikhman and A.V. Skorokhod. *Introduction to the Theory of Random Processes.* Saunders Company, 1969.

[35] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, 1987.

[36] E. Wong and B. Hajek. *Stochastic Processes in Engineering Systems.* Springer-Verlag, New York Berlin Heidelberg Tokyo, 1985.

APPENDIX C

Conditional-Mean Estimation Via Jump-Diffusion Processes in Multiple Target Tracking/Recognition *

M. I. Miller† and A. Srivastava† and U. Grenander‡

Abstract

A new algorithm is presented for generating the conditional mean estimates of functions of target positions, orientation and type in recognition and tracking of an unknown number of targets and target types. Taking a Bayesian approach a posterior measure is defined on the tracking/target parameter space by combining the narrowband sensor array manifold model with a high resolution imaging model, and a prior based on airplane dynamics. The Newtonian force equations governing rigid body dynamics are utilized to form the prior density on airplane motion. The conditional mean estimates are generated using a random sampling algorithm based on *Jump-Diffusion* processes, [1], for empirically generating MMSE estimates of functions of these random target positions, orientations and type under the posterior measure. Results are presented on target tracking and identification from an implementation of the algorithm on a networked Silicon Graphics and DECmpp/MasPar parallel machines.

*Submitted to the *IEEE Transactions on Signal Processing.* This paper was presented in part at the 29th Annual Allerton Conference on Communication, Control and Computing, 1991, University of Illinois at Urbana-Champaign, Illinois, and at the Communications and Informations Sciences and Systems Conference at Johns Hopkins University, March 1993. This work was supported by the ARO DAAL03-92-G-0141, ONR N00014-92-J-1418, ONR N00014-94-1-0859, and Rome Laboratory F30602-92-C-0004 to Michael I. Miller and ARO/MIT DAAL03-92-G-0115, ONR N00014-91-J-1021, ARL MDA972-93-1-0012, and NSF DMS-9217655 to Ulf Grenander.

†Department of Electrical Engineering, Electronic Signals and Systems Research Laboratory, Washington University St. Louis, MO. 63130

‡Division of Applied Mathematics, Brown University, Providence, Rhode Island

1 Introduction

This paper focuses on automated tracking and recognition of objects in remotely sensed complex dynamically changing scenes. Grenander's global shape models are used herein, extended to parametric representations of arbitrary and unknown model order, in which typical shape is represented via templates, with variability represented via transformation groups applied to the templates. The types of variability associated with the classical geometry are accommodated via the *Euclidean groups* involving both the rigid motions of translation and rotation. Since the objects are under dynamic motion, *the parameter spaces involves Cartesian products of these similarity groups.*

The second fundamental type of variability is associated with the model order (parametric dimension) and model type (recognition). In any scene there may be variable numbers of and different kinds of targets existing in the scenes for varying periods of time, implying the target number and therefore parametric dimension are unknown apriori. *Hence, the inference or hypothesis space becomes a search across countable disconnected unions of these Cartesian product groups*, with the model order and model type a variable to be inferred. We take a Bayesian approach, i.e. we define a prior distribution supported on this countable union of spaces, from which the posterior distribution is constructed. The parametric representation of the target scene is selected to correspond to *conditional expectations* under this posterior.

As we are particularly interested in non-cooperative moving targets, the algorithms are made robust to motion by incorporation of knowledge about motion dynamics into the prior distribution. The Newtonian force equations, a system of differential equations governing the motion of targets are used to induce the prior. These differential equations are parameterized by the target and or sensor type, and its orientation motion described by rotations in the special orthogonal group $SO(3)$ of 3×3 orthogonal matrices with determinant 1. It is the introduction

of these Newtonian force equations which makes tracking and recognition inseparable, since the equations of motion are explicitly parameterized by the sequence of airplane orientations. This provides the significant link between tracking algorithms based on data from narrowband sensors arrays in which the target is unresolved in the data (effectively a point), and high resolution information perhaps provided by a second sensor preserving the orientation information from which target recognition is performed. In part, it is this fundamental link which has motivated us to solve the tracking/recognition problem in a single consistent estimation framework in which the inference proceeds via the fusion of multi-sensor data: in our case, a *narrowband sensor array output* and *high-resolution images.*

Concerning the generation of conditional expectations, except under the most simplifying set of assumptions, the posterior distribution will be highly nonlinear in the parameters of hypothesis space, thus, precluding the direct closed form analytic generation of conditional expectations. Towards this end we have taken advantage of the explosion which has occurred over the past 10 years in the statistics community on the introduction of random sampling methods for the empirical generation of estimates from complicated distributions; see for example the reviews [2, 3]. Motivated by such approaches, we have previously described a new family of random sampling algorithms [4, 1] for generating conditional expectations in such disconnected hypothesis spaces. The random samples are generated via the direct simulation of a Markov process whose state moves through the hypothesis space with the *ergodic property* that the transition distribution of the Markov process converges to the posterior distribution. This allows for the empirical generation of conditional expectations under the posterior. To accommodate the connected and disconnected nature of the state spaces, the Markov process is forced to satisfy *jump-diffusion dynamics.* i.e. through the connected parts of the parameter space (Lie manifolds) the algorithm searches continuously, with sample paths corresponding to solutions of

standard diffusion equations; across the disconnected parts of parameter space the jump process determines the dynamics. The infinitesimal properties of these jump-diffusion processes are selected so that various sample statistics converge to their expectation under the posterior.

The original motivation for introducing jump-diffusions in [4, 1] is to accommodate the very different continuous and discrete components of the object discovery process. Given a conformation associated with a target type, or group of targets, the problem is to identify the orientation and translation parameters accommodating the variability manifest in the viewing of each object type. For this, the parameter space is sampled using diffusion search in which the state vector winds continuously through the similarities following gradients of the posterior. The second distinct part of the sampling process corresponds to the target type and number deduction during which the target types are being discovered, with some subset of the scene only partially "recognized" at any particular time during the process. The second type of change in parameter space are associated with a set of non-continuous transformations of the scene controlled by the jump process. A jump in hypothesis space corresponds to (i) jumping between different object types, (ii) hypothesizing a new object in the scene or a "change of mind" via the deletion of an object in the scene, or (iii) the merging or splitting of tracks and objects. The jump intensities are governed by the posterior density, with the process visiting configurations of higher probability for longer exponential times, and the diffusion equation governing the dynamics between jumps. It is the fundamental difference between diffusions (almost surely continuous sample paths) and jump processes (making large moves in parameter space in small time) which allows us to explore the very different connected and non-connected nature of hypothesis space.

Now automated target tracking and recognition are well known problems in the signal processing and control system's literature, with a great deal of published work on multiple target

tracking posed as state estimation problems [5, 6, 7]. In such approaches Kalman filter based techniques are emphasized, with linear descriptions of state playing a fundamental role. For situations in which the observed data are non-linear in target parameters the use of the extended Kalman filter has been proposed corresponding to linear approximations which prove valid for particular scenarios. There also now exists a substantial body of important work in tracking the directions of arriving signals from multiple moving sources recorded via sensor arrays [8, 9, 10]. In such sensor array based approaches the non-linear relationship between the parameters of motion and the sensor data are addressed directly, the linear Kalman filter state equations for tracking guiding or providing initial conditions for the gradient based estimators generated from the likelihood. In these non-linear data models, several variations of the gradient based techniques are used to solve the problem in mostly maximum-likelihood settings. However, the majority of researchers utilize simplifying assumptions which are not always valid in a general tracking scenario. For example, targets may be assumed stationary between sample times with multiple ($\sim$ 100) snapshots at each sample time, whereas, in general, for a moving target, each data sample reflects a new position. Also, though researchers base their models on simplified versions of target dynamics for the tracking scenario, mostly constant velocity - constant acceleration state constraint equations have been used because of their linear nature. These restricted motions are partly due to assumptions required for Kalman updating, but perhaps more fundamentally due to the separation of the tracking and recognition problems. The more informative priors used in this paper require high resolution recognition as the priors are coupled to the target type and its orientations. In part, this is one of the major results of this work.

In the work presented here, we define a random sampling based solution for generating minimum mean squared error estimates of the state variables for tracking and recognition problems in

a general setting. We assume data from a narrowband sensor array providing azimuth-elevation data for object tracking, and optical or radar imagers providing detailed information about the target-type and orientation. The goal is to track and recognize the unknown number of non-cooperative sources. The paper is organized as follows. In section 2 we define the parameter spaces with the posterior distribution derived in section 3. Section 4 describes an inference algorithm based on jump-diffusion processes and section 5 presents various results.

2 Recognition Via Deformable Templates

We use the global shape models and pattern theoretic approach introduced by Grenander [11, 12] to analyze complex scenes. As the basic building blocks of the hypotheses we define a subset of generators $\mathcal{G}^o$, which contains each target type $a \in \mathcal{A}$ ($\mathcal{A}$ is the alphabet of target types) placed at the origin of the inertial reference frame aligned to the inertial axes. The fundamental variability in target space is accommodated by applying the transformations $T(\phi)$ and $T(p)$ to the templates $g^o \in \mathcal{G}^o$ according to

$$T(\phi) : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi_1 & sin\phi_1 \\ 0 & -sin\phi_1 & cos\phi_1 \end{bmatrix} \begin{bmatrix} cos\phi_2 & 0 & -sin\phi_2 \\ 0 & 1 & 0 \\ sin\phi_2 & 0 & cos\phi_2 \end{bmatrix} \times \begin{bmatrix} cos\phi_3 & sin\phi_3 & 0 \\ -sin\phi_3 & cos\phi_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (1)$$

$$T(p) : \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \rightarrow \begin{bmatrix} x_1 + p_1 \\ x_2 + p_2 \\ x_3 + p_3 \end{bmatrix}, \quad (2)$$

where $\phi \in [0, 2\pi]^3$ with $0, 2\pi$ identified (herein referred to as the 3-dimensional torus $\mathcal{T}(3)$), and $p \in R^3$ is the translation vector. These parameterized transformations operate on the templates from $\mathcal{G}^o$ generating the full set of possible elements constituting any scene. The left panel of Figure 1 shows a rendering of one of the 3-D ideal targets $g^o \in \mathcal{G}^o$ under one such transformation.

The Bayes posterior is parameterized via the set of transformations, as well as the airplane type. A pattern consisting of a single track arises from a single target appearing and disappearing at random times $t_1^{(m)}, t_2^{(m)} \in [t_0, t]$ the observation period, with the m-th track parameter vector $x^{(m)}$ an element of the space $x^{(m)} \in (\mathcal{X}_0 \bigcup \sharp)^{[t_0,t]} \times \mathcal{A}$, $\mathcal{X}_0 \equiv \mathcal{T}(3) \times \Re^3$. The symbol $\sharp$ is used to denote the absence of the target from the scene. It will be useful for us to introduce the notation $x^{(m)}(\tau), \tau \in [t_0, t]$ to denote the set of parameters encoding the m-th target at time τ. An M-track parameter vector $x(M)$ becomes

$$x(M) \in \mathcal{X}_t(M) \equiv \left[\left(\mathcal{X}_0 \bigcup \sharp\right)^{[t_0,t]} \times \mathcal{A}\right]^M . \tag{3}$$

Since the number of the targets M is unknown a priori, the complete parameter space is defined as $\mathcal{X}_t = \bigcup_{M=0}^{\infty} \mathcal{X}_t(M)$. The estimation problem is to estimate the individual configurations as well as the number M.

3 Bayesian Posterior

Minimum mean squared error (*MMSE*) parameter estimates are generated via their empirical computation under the posterior measure. As the posterior is proportional to the product of the prior density and the observed data likelihood we first derive a prior on the parameter space followed by a model for the data generation which determines the posterior. For real

time estimation problems, the posterior density is an explicit function of t denoted $\pi_t(\cdot)$. In the Bayesian approach the estimates are for each time t conditioned on the data observed up to that time t.

3.1 Prior Density on Parameter Space $\mathcal{X}_t$

Airplane dynamics: The formulation of the prior measure on airplane positions is based on equations of motion for rigid bodies. We use an approach, in which the prior is induced via partial differential equations by assuming the forcing function to be a white process, which induces a Gaussian process with covariance corresponding to the differential operator expressing airplane dynamics. For this purpose, we use a formulation of airplane dynamics through differential equations as described by Cutaia and O'Sullivan [13]. Airplane dynamics are most straightforwardly expressed using the velocities projected along the body-fixed axes, called the body-frame velocities and here denoted $v(s) = [v_1(s)\ v_2(s)\ v_3(s)]$. They are depicted in the right panel of Figure 1.

Following standard rigid body analysis (see [14], for example) and neglecting the earth's curvature, motion and wind effects, the translational velocities $v(s)$ and rotational velocities $q(s) = [q_1(s)\ q_2(s)\ q_3(s)]$ satisfy the following set of differential equations:

$$\begin{aligned}
\dot{v}_1(s) - q_3(s)v_2(s) + q_2(s)v_3(s) &= f_1(s)\,, \\
\dot{v}_2(s) + q_3(s)v_1(s) - q_1(s)v_3(s) &= f_2(s)\,, \\
\dot{v}_3(s) - q_2(s)v_1(s) + q_1(s)v_2(s) &= f_3(s)\,, \\
I_1\dot{q}_1(s) - (I_2 - I_3)q_2(s)q_3(s) &= \Gamma_1(s)\,, \\
I_2\dot{q}_2(s) - (I_3 - I_1)q_1(s)q_3(s) &= \Gamma_2(s)\,, \\
I_3\dot{q}_3(s) - (I_1 - I_2)q_2(s)q_1(s) &= \Gamma_3(s)\,,
\end{aligned} \tag{4}$$

where $[f_1(s)\ f_2(s)\ f_3(s)]$ is the vector of applied translational forces, $[I_1\ I_2\ I_3]$ is the vector of rotational inertias, and $[\Gamma_1(s)\ \Gamma_2(s)\ \Gamma_3(s)]$ is the vector of applied torques. The first three equations describe the airplane's translational motion, while the next three describe its rotational motion.

At this time the prior which we have used for tracking is "somewhat less informative" in that only the first three equations, on translational motion, are used; detailed models of the targets associated with the torques for describing the rotational motion are not yet explicitly incorporated. The system matrix $A(\phi(s), \dot{\phi}(s))$ parameterizing Eqns. 4 is

$$\begin{bmatrix} 0 & -q_3(s) & q_2(s) \\ q_3(s) & 0 & -q_1(s) \\ -q_2(s) & q_1(s) & 0 \end{bmatrix},$$

with the velocities, inertial positions and Euler angles are related using the standard transformation to relate body-frame velocities with inertial frame positions according to

$$p(s) = \int_{t_0}^{s} \Psi(\tau) v(\tau) d\tau + p(t_0), \tag{5}$$

where $p(t_0)$ the initial position is assumed known and $\Psi(\tau)$ is the standard orthogonal rotation matrix given in Eqn. 1 parameterized by the Euler angles $\phi(\tau)$. The rotational motion determines the prior since with reference to the fixed inertial frame the angular velocity projections, $q(s)$, onto the rotating body axes determine the system matrix $A(\phi(s), \dot{\phi}(s))$. The $q(s)$ vectors are none other than the rates of change of the Euler orientation angles according to $q_1 = \dot{\phi}_1 - \dot{\phi}_3 sin(\phi_2),\ q_2 = \dot{\phi}_2 cos(\phi_1) + \dot{\phi}_3 cos(\phi_2) sin(\phi_1),\ q_3 = -\dot{\phi}_2 sin(\phi_1) + \dot{\phi}_3 cos(\phi_2) cos(\phi_1)$.

For the construction of the "informative" part of the prior, first condition the linear dif-

ferential equations on the sequence of system matrices $A(\phi(s), \dot{\phi}(s))$, via conditioning on the sequence of Euler rotation angles. Then the velocity process is a conditional Gaussian process induced by assuming the forcing function on the momentum equation to be a white process of fixed spectral density. The covariance function is derived as follows. Define the state transition matrix $\Phi(\tau, \cdot)$ as the unique solution of the matrix differential equation

$$\frac{dM(s)}{ds} = -A(\phi(s), \dot{\phi}(s))M(s) \; , \quad M(\tau) = I \; , \tag{6}$$

then the covariance of the body frame velocity process becomes

$$\mathcal{K}_v(s_1, s_2) = \sigma \int_{t_0}^{min(s_1,s_2)} \Phi(t_1, s_1)\Phi^{\dagger}(t_1, s_2)dt_1 + \Phi(t_0, s_1)\mathcal{K}_v(t_0, t_0)\Phi^{\dagger}(t_0, s_2) \; , \tag{7}$$

where $\mathcal{K}_v(t_0, t_0)$ is the covariance of the initial velocity, $v(t_0)$. The inertial position process is then Gaussian with covariance $\mathcal{K}_p(s_1, s_2) = \int_{t_0}^{s_1} \int_{t_0}^{s_2} \Psi(\tau_1)\mathcal{K}_v(\tau_1, \tau_2)\Psi^{\dagger}(\tau_2)d\tau_1 d\tau_2$. The covariance function is parameterized by the sequence of airplane orientations thereby demonstrating the fundamental link between tracking unresolved targets and high-resolution recognition algorithms.

The more "diffuse" component of the prior is developed by assuming the Euler angles are fixed for small sampling intervals, giving a sequence of angles $\phi(1), \phi(2), \ldots, \phi(j) = \phi(s), s \in [j\Delta, (j+1)\Delta)$. Then, the marginal on $\phi(j)$ takes the form of a Markov Von-Mises prior on the torus $\mathcal{T}(3)$ (see e.g. [15]) with the density $\prod_{i=1}^{3} \frac{1}{2\pi I_0(\kappa_i)} e^{\kappa_i cos(\phi_i(j) - \bar{\phi}_i(j))}$, where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero, and $\kappa = [\kappa_1 \; \kappa_2 \; \kappa_3]$ is the vector of concentration parameters, and $\bar{\phi(j)} = [\bar{\phi}_1(j) \; \bar{\phi}_2(j) \; \bar{\phi}_3(j)]$ is the mean of $\phi(j)$. The orientation process is made Markov by assigning the previous state as the mean of the present state, giving

a potential of the form

$$\sum_j \sum_{i=1}^{3} \kappa_i cos(\phi_i(j) - \phi_i(j-1)) \; . \tag{8}$$

Recognition: We want to drive the algorithm towards deductions which are as simple as possible. Therefore, we use priors based on run-length coding to encourage hypotheses with minimal numbers of aggregated tracks. For this, associate with a target appearing at time $t_1^{(m)}$ and exiting at time $t_1^{(m)} + t_2^{(m)}$ the number of bits $log^* t_1^{(m)} + log^* t_2^{(m)} + log|\mathcal{A}| + \frac{6}{2} t_2^{(m)} log(sample - size)$, log^* the iterated logarithm $log + loglog + \ldots$ defined by Rissanen [16, 17] for constructing priors on the reals. Then the *complexity* prior for an M-track scene has potential

$$\left(log^* t_1^{(m)} + log^* t_2^{(m)} + log|\mathcal{A}| + \frac{6}{2} t_2^{(m)} log(sample - size) \right) \; . \tag{9}$$

3.2 Data Likelihood

The likelihood of the collected data correspond to two sensor types, a *tracking sensor* consisting of an array of passive sensors and a range radar, and a *high-resolution imaging* sensor.

Low resolution tracking: As shown in the left panel of Figure 2, for azimuth-elevation coordinate tracking, a cross array of n isotropic sensors is assumed as in [18, 19, 20, 21] using the standard narrowband signal model developed in [22]. Accordingly, depending on the geometry of the sensor arrangement, the phase lags of the signal reaching different sensor elements are known functions of the source locations. The deterministic signal model for sensor arrays is used in which the $n \times 1$ sensor array measurement vector $y_1(\tau)$, $\tau \in [t_0, t]$ is complex Gaussian distributed with diagonal covariance and mean $E\{y_1(\tau)\} = \sum_{m=1}^{\infty} d(x^{(m)}(\tau)) 1_{\mathcal{X}_0}(x^{(m)}(\tau)) s^{(m)}(\tau)$, with $s^{(m)}(\tau)$ the signal amplitude of the m-th track at time τ. Notice the indicator function $1_{\mathcal{X}_0}(x^{(m)}(\tau))$ selects the targets that contribute to the array manifold at time τ, $d(p^{(m)}(\tau))$ is the direction vector determined by the array geometry and the position of the m-th target.

Since the tracking array responds to the inertial positions of the target most naturally in azimuth and elevation, we convert from rectangular coordinates $p(t) = [p_1(t)\ p_2(t)\ p_3(t)] \in R^3$ to polar coordinates, $[r(t)\ \alpha_1(t)\ \alpha_2(t)] \in R^+ \times [0, 2\pi)^2$, (range, elevation and azimuth) using the standard relationship,

$$r = \sqrt{p_x^2 + p_y^2 + p_z^2}\ , \alpha_1 = arctan\frac{p_z}{\sqrt{p_x^2 + p_y^2}}\ , \alpha_2 = arctan\frac{p_y}{p_x}\ . \tag{10}$$

High resolution imaging: While the statistical models for high-resolution radar imaging are being incorporated in this problem by others ([23, 24, 25, 26, 27, 28]), all of the results shown here are based on an optical imaging system as depicted in the right panel of Figure 2. In this system, the data are a sequence of 2-D images resulting from projecting the scene volume containing targets onto the focal plane of the imaging sensor; i.e., the imaging process is modeled as a far field orthographic projection. Since the parameter set $x(M)$ completely determines the imaged volume, the projection is a deterministic operation from the parameter space to the measurement space $\Re^{\mathcal{L}\times[t_0,t]}$, $\mathcal{P} : \mathcal{X}_t \rightarrow \Re^{\mathcal{L}\times[t_0,t]}$, where $\mathcal{L}$ is the 2-D image space. For all of the results shown here the high-resolution imaging data is a non-zero mean white Gaussian process with mean the projective transformation of the scene: $E\{y_2(\tau)\} = \mathcal{P}(x(\tau)),\ \tau \in [t_0, t]$.

The posterior distribution is obtained as the product of the data likelihood and the prior density and is defined explicitly by Eqns. 11,12 below.

Remark: For observing the range locations of the targets, a range radar is assumed with the observations modeled as normally distributed with mean $|p(\tau)|$, the 2-norm of the position vector at time $\tau \in [t_0, t]$.

4 Random Sampling for Generating Conditional Expectations

4.1 The parameter space.

In particularizing the jump-diffusion algorithm to the multiple tracking problem it will be convenient to suppress explicit dependence on time t. Therefore, for each time t we will have a distribution for which the jump-diffusion process will be constructed. The parameter spaces themselves will be indexed by t, and form an increasing family of spaces.

The crucial part of the problem still remaining is the derivation of the inference algorithm: For all of the possible scenes we assume that the targets are stationary during some fundamental data sampling intervals, with the parameters and sensor data represented by their values on some index set $\{\tau_j\}_{j=1,\ldots,J}$, J the total number of sample points in the observation interval $\tau_j \in [t_0, t]$. Note, J is actually a function of t. Then an M-track parameter vector becomes $x(M) \in \mathcal{X}(M) \equiv (\mathcal{X}_0 \bigcup \{\sharp\})^{MJ} \times \mathcal{A}^M$ with the complete parameter space being $\mathcal{X} = \bigcup_{M=0}^{\infty} \mathcal{X}(M)$. It will be useful to define the number of track segments in the m-th track $n^{(m)}$, and the total number of track-segments as $n(M)$, implying for example, $x(M) \in \mathcal{X}_0^{n(M)} \times \mathcal{A}^M$.

The **posterior** μ is of the Gibb's form with the potential H_M for $x(M) \in \mathcal{X}(M)$ becoming

$$\begin{aligned} H_M(x(M)) \;=\; & 1/\sigma_1 \sum_{j=1}^{J} |y_1(\tau_j) - \sum_{m=1}^{M} d(x^{(m)}(\tau_j)) 1_{\mathcal{X}_0}(x^{(m)}(\tau_j))|^2 + 1/\sigma_2 \sum_{j=1}^{J} |y_2(\tau_j) - \mathcal{P}(x(\tau_j))|^2 \\ & + \sum_{m=1}^{M} \left(p^{(m)} K_p^{(m)} p^{(m)} + \sum_{j=2}^{n^{(m)}} \sum_{i=1}^{3} \kappa_i^{(m)} cos(\phi_i^{(m)}(\tau_j) - \phi_i^{(m)}(\tau_{j-1})) \right) \\ & + \sum_{m=1}^{M} \left(log^* n^{(m)} + log^* t_1^{(m)} + log|\mathcal{A}| + \frac{6}{2} t_2^{(m)} log(sample - size) \right) . \end{aligned} \tag{11}$$

The first two quadratic terms are associated with the tracking data and the high resolution imaging data. The last three terms are the prior terms on the tracking parameters, Von-Mises orientations and track complexity, respectively For arbitrary $x \in \mathcal{X}$, define the potential

$H_M(x) = 0$, for $x \notin \mathcal{X}(M)$. Then the posterior measure $\mu(\cdot)$ with density $\pi(\cdot)$ is in Gibb's form according to

$$\mu(dx) = \frac{\sum_{M=0}^{\infty} e^{-H_M(x)} 1_{\mathcal{X}(M)}(x)}{\mathcal{Z}} dx \ , \tag{12}$$

with the normalizer $\mathcal{Z} = \sum_{M=0}^{\infty} \int_{\mathcal{X}(M)} e^{-H_M(x)} dx$.

Now for the development which follows it will be convenient to define the part of the posterior which does not include the prior. This we define as L_M and is given by the formula

$$L_M(x(M)) = 1/\sigma_1 \sum_{j=1}^{J} |y_1(\tau_j) - \sum_{m=1}^{M} d(x^{(m)}(\tau_j)) 1_{\mathcal{X}_0}(x^{(m)}(\tau_j))|^2 + 1/\sigma_2 \sum_{j=1}^{J} |y_2(\tau_j) - \mathcal{P}(x(\tau_j))|^2 \ ,$$

with the prior potential term denoted by

$$\begin{aligned} P_M(x(M)) \ = \ & \sum_{m=1}^{M} \left(p^{(m)} K_p^{(m)} p^{(m)} + \sum_{j=2}^{n^{(m)}} \sum_{i=1}^{3} \kappa_i^{(m)} cos(\phi_i^{(m)}(\tau_j) - \phi_i^{(m)}(\tau_{j-1})) \right) \\ & + \sum_{m=1}^{M} \left(log^* n^{(m)} + log^* t_1^{(m)} + log|\mathcal{A}| + \frac{6}{2} t_2^{(m)} log(sample - size) \right) \ . \end{aligned}$$

Remark: Notice, in identifying model M with an M-track configuration the potential must be adjusted so that the m-th track covariance $K_p^{(m)}$ is an $n^{(m)} \times n^{(m)}$ matrix appropriate for the m-th track. We would be more precise by identifying models $(n^{(1)}, n^{(2)}, \ldots n^{(M)}, M)$ with index $k \in \aleph$, from which the potential is then uniquely defined in the usual sense. However, with the subtle breach of notational convention we simply define H_M as the potential associated with an M-track configuration and modify the potential according to the variable numbers of parameters associated with the tracks.

4.2 The basic jump-diffusion set-up.

The crucial part of the problem still remaining is the derivation of the inference algorithm: that is *how shall we carry out hypothesis formation?* We follow the analysis outlined in [4, 1], in which conditional expectations with respect to the posterior density $\pi(\cdot)$ are generated empirically. First identify with each model an index $k \in \aleph$ with parameter space $\mathcal{X}(k)$ of dimension $n(k)$. The full hypothesis space $\mathcal{X} = \cup_{k=0}^{\infty} \mathcal{X}(k)$. The posterior distribution μ is then of the Gibb's type supported on $\mathcal{X}$, i.e. for all set $\mathcal{A} \subset \mathcal{X}$ Lebesgue measurable,

$$\begin{aligned} \mu(\mathcal{A}) &= \sum_{k=0}^{\infty} \mu(\mathcal{A} \cap \mathcal{X}(k)) \, , \\ &= \sum_{k=0}^{\infty} \int_{\mathcal{A} \cap \mathcal{X}(k)} \frac{e^{-H_k(x(k))}}{\mathcal{Z}} dx(k) \, . \end{aligned} \tag{13}$$

The goal is to essentially sample from μ generating a sequence of samples $X(s_1), X(s_2), \ldots$ with the property that

$$1/n \sum_{j=1}^{n} f(X(s_j)) \stackrel{n \to \infty}{\to} \int_{\mathcal{X}} f(x) \mu(dx) \, . \tag{14}$$

This we do via the construction of a Markov process $X(s)$ which satisfies *jump-diffusion dynamics* through $\mathcal{X}$ in the sense that (i) on random exponential times the process jumps from one of the countably infinite set of spaces $\mathcal{X}(k), k = 0, 1, \ldots$ to another, and (ii) between jumps it satisfies diffusions of dimension $n(k)$ appropriate for that space.

The process $X(s)$ within each of the multiple sub-spaces is a diffusion with infinitesimal drifts $a(x(k)) \in R^{n(k)}$ and infinitesimal variance matrix $B(x(k))$, an $n(k) \times n(k)$ matrix. It is the existence of this multiple disconnected union of spaces $\mathcal{X}$ which motivates the introduction of the second transformation type on the models, transformations which act by changing one model type to another with its resulting configuration. The transformations we shall term *simple moves*

which are drawn probabilistically from a family $\mathcal{F}$ of changes in the model types $k \in \aleph$, and are applied discontinuously, with the simple moves defining transitions through $\aleph$, $\mathcal{F} : \aleph \rightarrow \aleph$. The family of transitions are chosen large enough to act transitively in the sense that given any pair $k', k'' \in \aleph$ it should be possible to find a finite chain of transitions that leads from k' to k''.

The set $\mathcal{F}$ controls the jump dynamics in the jump-diffusion processes as follows. The jump process corresponds to movement from one subspace to another on the jump times with transition probability measure $Q(x, dy) = \frac{q(x,dy)}{q(x)}$, $\int_{\mathcal{X}} Q(x, dy) = 1$. The measures $q(x, dy)$ are defined in the standard way [29]: $q(x, dy) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon}\Big(\Pr\{X(s+\epsilon) \in dy | X(s) = x\} - 1_{dy}(x)\Big)$, with $q(x) = \int_{\mathcal{X} \setminus x} q(x, dy)$. The set $\mathcal{F}$ determines which measures are non-zero.

As we have shown for purely Euclidean spaces [4, 1], the proper choice of jump transition measures and diffusion drifts and variances (stochastic gradients) will make μ on $\mathcal{X}$ invariant with ergodic averages generated from the process converging to their expectations. See theorem 1 of [1].

4.2.1 The jump process.

The **jump process** is controlled by the family of changes $\mathcal{F}$ which control the movement through the non-connected subspaces. Changes in model type will include increasing and decreasing track length, increasing and decreasing number of tracks, and changing the target types. The set of transformations are defined as $\mathcal{F} = \{\vartheta^d_{t(j)}, \vartheta^d_{s(j)}, \vartheta^b_{t(j)}, \vartheta^b_{s(j)}, \vartheta_{a(j)}\}$. The first two correspond to deletion or removal of the j-th track and segment, which are mappings $\vartheta^d_{t(j)} : \mathcal{X}_0^{n(M)} \times \mathcal{A}^M \rightarrow \mathcal{X}_0^{n(M)-1} \times \mathcal{A}^{M-1}$, $\vartheta^d_{s(j)} : \mathcal{X}_0^{n(M)} \times \mathcal{A}^M \rightarrow \mathcal{X}_0^{n(M)-1} \times \mathcal{A}^M$, respectively. The second two are birth operators birthing tracks and segments to the j-th place or track, and are mappings $\vartheta^b_{t(j)} : \mathcal{X}_0^{n(M)} \times \mathcal{A}^M \rightarrow \mathcal{X}_0^{n(M)+1} \times \mathcal{A}^{M+1}$, $\vartheta^b_{s(j)} : \mathcal{X}_0^{n(M)} \times \mathcal{A}^M \rightarrow \mathcal{X}_0^{n(M)+1} \times \mathcal{A}^M$, respectively. The last operator simply changes the target type, $\vartheta_{a(j)} : \mathcal{X}_0^{n(M)} \times \mathcal{A}^M \rightarrow \mathcal{X}_0^{n(M)} \times \mathcal{A}^M$. It should

be noted that the addition of only unit length tracks is allowed, and unit length track segments, as well as deletions of only unit length tracks or segments. Define the set, of indices of tracks in $x(M)$ which are candidates for deletion, by $\{m : 1 \leq m \leq M, n^{(m)} = 1\}$ and let $M_1(x(M))$ be the cardinality of this set. For the increments in parameter space, an explicit notation denoting a specific segment or track added to the configuration will be needed. Let $\oplus_j$ stand for the addition of track segments or tracks to the existing configuration, i.e. $x(M) \oplus_j y(1)$ represents an $M+1$ track configuration formed by adding $y(1)$ to $x(M)$ at the j-th location in the list, and $x(M) \oplus_j y$ signifies addition of a segment to the j-th track of $x(M)$.

These are the only transformations of model type that are allowed. To carry the evolution of the state forward from the diffusion we make the jump measures singular with respect to the Lebesgue measures in the respective subspaces which the jump transformations move into. For this, the part of the state which is not being added or deleted remains unchanged after the jump transformation. This corresponds to the following transition measures of the type,

$$
\begin{aligned}
q(x(M), dy(M+1)) &= \sum_{j=1}^{M+1} q_t^b(x(M), y(M+1))\delta_{x(M)}\Big(d(\vartheta^d_{t(j)} y(M+1))\Big)dy(1)\,, \\
q(x(M), dy(M)) &= \sum_{j=1}^{M} q_s^b(x(M), y(M))\delta_{x(M)}\Big(d\Big(\vartheta^d_{s(j)} y(M)\Big)\Big)dy \\
&+ \sum_{j=1}^{M} q_s^d(x(M), y(M))\delta_{\vartheta^d_{s(j)} x(M)}(dy(M))\,, \\
&+ \sum_{j=1}^{M} q_a(x(M), y(M))\delta_{x(M)}\Big(d\Big(\vartheta_{a(j)} y(M)\Big)\Big)\,, \\
q(x(M), dy(M-1)) &= \sum_{j=1}^{M} q_t^d(x(M), y(M-1))\delta_{\vartheta^d_{t(j)} x(M)}(dy(M-1)) \qquad (15)
\end{aligned}
$$

Let $\mathcal{F}^1(x(M)) \subset \aleph$ be the set of models that can be reached from $x(M)$ in one jump move, and $\mathcal{X}(\mathcal{F}^1(x(M)))$ the space containing the configurations of these types. The total jump

intensity becomes

$$q(x(M)) = \int_{\mathcal{X}(\mathcal{F}^1(x(M)))} q(x(M), dy) \, . \tag{16}$$

4.2.2 The diffusion process.

The **diffusion process** between jumps controls the dynamics of $X(s)$ in their respective sub-spaces. For the sub-space associated with M tracks having $n(M)$ segments the diffusion flows through the manifold $\mathcal{X}_0^{n(M)} = \mathcal{T}(3)^{n(M)} \times R^{3n(M)}$ associated with the orientations $\phi \in \mathcal{T}(3)$ and positions $p \in R^3$. The restriction of a previous result in Theorem 1 from [1] to Euclidean spaces unfortunately prevents its direct application to the tracking problem in which the torus is involved. Even the most innocuous appearing stochastic differential equation (S.D.E.) make little sense when the manifold is curved in any way. This forces us to use more general results on Lie manifolds as described in [30] and adapted to the tracking case as follows.

Associate with the first $3n(M)$ components of the state vector the flow through $\mathcal{T}(3)^{n(M)}$, and the last $3n(M)$ components the flow through $R^{3n(M)}$ according to $X(s) = [X_1(s), X_2(s)]$, $X_1(s) \in \mathcal{T}(3)^{n(M)}$, $X_2(s) \in R^{3n(M)}$. Also, define $\frac{e^{-P(y(1)|x(M))}}{Z_T(1)}$, $\frac{e^{-P_j(y|x(M))}}{Z_S(1)}$ as the conditional prior densities on the single track and single segment spaces given the current configuration $x(M)$, respectively, with $Z_T(1), Z_S(1)$ being their normalizers. P_j denotes the attachment of the segment y to the j^{th}-track of the set $x(M)$. Then we have the following theorem.

Theorem 1 *IF the jump diffusion process $X(s)$ has the properties that*

1. *the diffusion $X(s)$ within any of the subspaces $\mathcal{X}_0^{n(M)}$ satisfies the S.D.E.*

$$X_1(s) = \left[X_1(0) + \int_0^s -\frac{1}{2} \nabla_1 H_M(X(\tau)) d\tau + W_1(s) \right]_{\text{mod } 2\pi} , \tag{17}$$

$$X_2(s) = X_2(0) + \int_0^s -\frac{1}{2} \nabla_2 H_M(X(\tau)) d\tau + W_2(s) \, , \tag{18}$$

where $[\cdot]_{\text{mod } 2\pi}$ *is taken componentwise,* ∇_1, ∇_2 *are the gradients with respect to the orientation and position (velocity) vectors respectively, and* $W_1(s), W_2(s)$ *are standard vector Wiener processes of dimensions* $3n(M)$,

2. *and the birth/death parameters of the jump measures*

$$
\begin{aligned}
q_t^b(x(M), x(M) \oplus_j y(1)) &= \frac{1}{5(M+1)} e^{-[L_{M+1}(x(M)\oplus_j y(1)) - L_M(x(M))]_+} \frac{e^{-P(y(1)|x(M))}}{Z_T(1)}, \quad j = 1, .., M+1 , \\
q_s^b(x(M), x(M) \oplus_j y) &= \frac{1}{5M} e^{-[L_M(x(M)\oplus_j y) - L_M(x(M))]_+} \frac{e^{-P_j(y|x(M))}}{Z_S(1)}, \quad j = 1, .., M , \\
q_t^d(x(M), \vartheta^d_{t(j)} x(M)) &= \frac{1}{5M_1(x(M))} \frac{e^{-[L_{M-1}(\vartheta^d_{t(j)} x(M)) - L_M(x(M))]_+}}{Z_T(1)} 1_{\{m>0\}}(M_1(x(M)) , \\
& \qquad\qquad j \in \{m : 1 \le m \le M, n^{(m)} = 1\} , \\
q_s^d(x(M), \vartheta^d_{s(j)} x(M)) &= \frac{1}{5M} \frac{e^{-[L_M(\vartheta^d_{s(j)} x(M)) - L_M(x(M))]_+}}{Z_S(1)} , \quad j = 1, .., M , \\
q_a(x(M), \vartheta_{a(j)} x(M)) &= \frac{1}{5M} e^{-[L_M(\vartheta_{a(j)} x(M)) - L_M(x(M))]_+}, \quad j = 1, .., M \quad . \qquad (19)
\end{aligned}
$$

THEN, $X(s_j)$ *converges in variation norm to* μ.

Proof: The proof follows the general approach in [1, 30] with details summarized for this problem in the Appendix.

4.3 Algorithm Implementation

The jump-diffusion process satisfying Theorem 1 is constructed as follows. Initialize with $t_0 = 0, i = 0$.

1. Generate an exponential random variable u with mean 1.

2. For $s \in [s_i, s_i + u)$, $X(s)$ follows the stochastic differential Eqns. 17,18 in subspace determined by $X(s_i)$.

3. On random time $s_{i+1} = s_i + u$, define $x_{\text{old}} = X(s_{i+1}^-)$ and determine $M = M_{\text{old}}$ of x_{old} the number of tracks in $X(s_{i+1}^-)$.

4. Draw one of the 5 possible jump choices from the set $\{t^b, s^b, t^d, s^d, a\}$ according to the distribution $\{\frac{1/5}{Z}, \frac{1/5}{Z}, \frac{1/5 Z_T(1)}{Z} 1_{\{m>0\}}(M_1(x_{old})), \frac{1/5 Z_S(1)}{Z}, \frac{1/5}{Z}\}$, with $Z = \frac{1}{5} + \frac{1}{5} Z_T(1) 1_{\{m>0\}}(M_1(x_{old})) + \frac{1}{5} Z_S(1) + \frac{1}{5} + \frac{1}{5}$.

   If, t^b, then draw a 1-length track $y(1)$ from a uniform prior on $\mathcal{X}_0 \times \mathcal{A}$ and draw $j \in \{1, 2, \ldots, M+1\}$ uniformly:

$$x_{\text{new}} \leftarrow x_{\text{old}} \oplus_j y(1) \ .$$

   Else If, t^d, draw $j \in \{m : 1 \leq m \leq M, n^{(m)} = 1\}$ uniformly:

$$x_{new} \leftarrow \vartheta^d_{t(j)} x_{\text{old}} \ .$$

   Else If, s^b, then draw $j \in \{1, 2, \ldots, M\}$ uniformly and draw $y \in \mathcal{X}_0$ from the Von-Mises prior on $\mathcal{T}(3)$ and the Gaussian prior on $\Re^3$, $\frac{e^{-P_j(y|x(M))}}{Z_S(1)}$, conditioned on the current j^{th} track configuration:

$$x_{new} \leftarrow x_{\text{old}} \oplus_j y \ .$$

   Else If, s^d, draw $j \in \{1, 2, \ldots, M\}$ uniformly:

$$x_{new} \leftarrow \vartheta^d_{s(j)} x_{\text{old}} \ .$$

   Else, Draw $j \in \{1, 2, \ldots, M\}$ uniformly:

$$x_{new} \leftarrow \vartheta_{a(j)} x_{\text{old}} \ .$$

5. Determine M_{new} of x_{new}.

6. If, $L_{M_{\text{old}}}(x_{\text{old}})) - L_{M_{\text{new}}}(x_{\text{new}}) > 0$, $X(s_{i+1}) \leftarrow x_{\text{new}}$

   Else $X(s_{i+1}) \leftarrow x_{\text{new}}$ with probability $e^{-[L_{M_{\text{new}}}(x_{\text{new}}) - L_{M_{\text{old}}}(x_{\text{old}})]}$ and

   Else $X(s_{i+1}) \leftarrow x_{\text{old}}$ with probability $1 - e^{-[L_{M_{\text{new}}}(x_{\text{new}}) - L_{M_{\text{old}}}(x_{\text{old}})]}$.

7. $i \leftarrow i + 1$, return to 1.

Since a track-segment of length 1 correspond to $y \in \mathcal{X}_0$ consisting of the position and orientation components, the discretized form of Eqn. 4 is used for the position and the Markov Von-Mises prior on the torus $\mathcal{T}(3)$ used for the orientation component. The candidates for deletion are obtained by removing the last segment from the current j^{th} track estimate or the j^{th} track estimate itself.

5 Results

Below we focus on single track identification in 3-D space; see [31] for multiple target tracking in 2-D.

For the implementation of the jump-diffusion algorithm for estimating the motion of a single target, i.e. M = 1, the parameter space becomes $\mathcal{X}_0^{[t_0,t]} \times \mathcal{A}$. For the implementation there are a total of two target types, $\mathcal{A} \equiv \{1, 2\}$. The algorithm was jointly implemented using the *flight simulator* software on the Silicon Graphics workstation for generating the data sets, and a massively parallel 4096 processor SIMD DECmpp/MasPar machine for implementing the tracking-recognition algorithm. Figure 4 shows the simulation environment for a sample target-flight observed by sensor systems located on ground represented by the mesh. The target motion is observed at 1500 times during the flight.

The array geometry corresponds to a 64-element cross-array of isotropic sensors located at

half-wavelength spacing. The tracking data $\{y_1(\tau), \tau \in [t_0, t)\}$ is a 64-element complex vector with mean $d(p(\tau))s(\tau)$ and additive complex Gaussian white noise of the Goodman's class. The direction vector corresponds to the array takes the form

$$[e^{-\frac{i31}{2}\lambda_1(\tau)},\ e^{-\frac{i29}{2}\lambda_1(\tau)}\ ,...,e^{\frac{i31}{2}\lambda_1(\tau)},\ e^{-\frac{i31}{2}\lambda_2(\tau)},e^{-\frac{i29}{2}\lambda_2(\tau)}\ ,..,\ e^{\frac{i31}{2}\lambda_2(\tau)}]^T\ ,\quad i=\sqrt{-1}\ ,$$

where $\lambda_1(\tau) = \pi cos(\alpha_1(\tau))sin(\alpha_2(\tau))$, and $\lambda_2(\tau) = \pi cos(\alpha_1(\tau))sin(\alpha_2(\tau))$, $\alpha_1(\tau), \alpha_2(\tau)$ is the azimuth, elevation angles of the target position $p(\tau)$. Since we use the velocity representation the azimuth and elevations are generated using the standard coordinate transformation of Eqns. 5,10. The upper panels in Figure 3 display the azimuth-elevation power spectra of the tracking data, generated by projecting the data vector onto the candidate direction vectors, for two target locations.

The 2-D imaging data $\{y_2(\tau), \tau \in [t_0, t)\}$ consists of 4096 Gaussian random variables associated with a 64×64 imaging lattice. The mean is $\mathcal{P}(x(t))$ where $\mathcal{P}(\cdot)$ is simply the 2-D projection of the rendered object positioned and oriented at $p(\tau), \phi(\tau)$, with additive noise. The lower panels in Figure 3 show two data samples obtained by high resolution imaging of the target along its flight.

At any given time t the jump-diffusion algorithm is run to generate samples from the posterior distribution generated by the data up to time t. This simulation is performed until the next data set arrives at $t + 1$ when the algorithm starts sampling from the new posterior. For sampling the jump-diffusion Markov process is constructed as follows. For the single object case the possible jump transformations through parameter space involve either addition of a track segment $y \in \mathcal{Y}_0$, deletion of a track segment, or a change of target type. The set of changes

$\mathcal{F} \equiv \{\vartheta^d_{s(1)}, \vartheta^b_{s(1)}, \vartheta_{a(1)}\}$ are transformations of the type

$$x(1) \in \mathcal{X}_0^{n(1)} \times \mathcal{A} \;\rightarrow\; x(1) \oplus_1 y \in \mathcal{X}_0^{n(1)+1} \times \mathcal{A}\,, \tag{20}$$

$$x(1) \in \mathcal{X}_0^{n(1)} \times \mathcal{A} \;\rightarrow\; \vartheta^d_{s(1)} x(1) \in \mathcal{X}_0^{n(1)-1} \times \mathcal{A}\,, \tag{21}$$

$$x(1) \in \mathcal{X}_0^{n(1)} \times \mathcal{A} \;\rightarrow\; \vartheta_{a(1)} x(1) \in \mathcal{X}_0^{n(1)} \times \mathcal{A}\,. \tag{22}$$

Shown in Figure 4 is the evolution of the random sampling algorithm for estimating the target track. The grey track represents the true airplane path, consisting of 1500 track segments, used in data generation with the estimated track shown overlapping in black at three different times during the estimation. Figure 5 shows a magnified view of a section of the track, formed of 8 track-segments, being estimated by the jump-diffusion algorithm. The top 4 panels illustrate the jump part of the algorithm for which we have turned off the diffusion. These upper panels shows successive guesses of the jump process which continually attempt to add and delete new track segments. Since the actual object has created a path which is longer then that which has been inferred by the algorithm during the early segments, the jump process always chooses to add new track segments. Notice, that on each addition the new segment is drawn from the prior on flight dynamics, which are parameterized by the track up to that point in time. Hence, the jump algorithm tends to infer track segments which are close to the true track if the current state vector is close to it. Because the diffusion has been turned off, notice the disparity between the track and the state of the algorithm.

The lower panels show the result of applying the diffusion to the state vector. The flow of the panels corresponds to increasing simulation time as the diffusion simulates from the posterior with the state brought into alignment.

Figure 6 depicts the importance of the dynamics based prior. Based on the equations of

motion and the track history the candidate segments are generated and accepted/rejected according to their likelihood. To show the support of the prior distribution in "phase space", the upper panels plot the 10 highest prior probability candidates placed at the track end for the algorithm to choose from. Each panel corresponds to a different time during the inference. The top row shows that if the track vector is close to the true track, the cone of candidates predicts well the future position. The lower panel shows the effect of the track state deviating from the true path, where the cone of prediction is not close to the future airplane position.

Figure 7 demonstrates the global importance of the prior distribution in estimating a portion of the target path. The algorithm was run with and without the prior measure, under the same parameters, with the results shown in the figure. The upper panels show the sequence of estimates obtained from the algorithm without any information from airplane dynamics. The lower panels use the prior information based on the equations of motion describing the airplane flight.

6 Acknowledgments

We would like to thank Robert Teichman for a great deal of help in the recognition component of the algorithm. We owe special thanks to Professor Yali Amit who provided the detailed development for the diffusion on Lie groups of which the torus is a particular example.

7 Appendix

Proof of Theorem 2: The proof has two parts: (i) showing that π is an invariant density of the process, and (ii) verifying that the process is irreducible and therefore π is the unique invariant density. Part (ii) follows directly that in [30] using the properties of the jump process and the

fact that the diffusions are each irreducible over their respective subspaces. In part (i) we need to verify the stationarity for both the jump and diffusion components of the Markov process. The generator, or backward Kolmogoroff operator, for the jump-diffusion process (denote it as $A = A^d + A^j$ (diffusion+jump)) characterizes the stationary density in that $\pi(x)$ is stationary for the jump-diffusion if and only if

$$\int Af(x)\pi(x)dx = 0 \tag{23}$$

for all f in the domain of A, $\mathcal{D}(A)$.

The diffusion process has two components corresponding to the S.D.E. on the multi-dimensional torus (17) and the S.D.E. on the Euclidean space of target positions (velocities) (18). To prove invariance of $\pi(x)$ for the diffusion on the torus we use results from [30] on invariant distributions of S.D.E.'s on manifolds, in particular the multi-dimensional torus. To demonstrate the approach, we prove the stationarity condition for the Euclidean component only. Define a set of functions which forms the domain of the generator A as

$$\mathcal{D}(A) = \{f : f = \sum_{M=0}^{K} 1_{\mathcal{X}(M)} f_M, f_M \in \hat{C}^2(\mathcal{X}(M)), K \geq 0\}\ , \tag{24}$$

with $\hat{C}^2$ twice continuously differentiable functions vanishing at ∞. Then the infinitesimal generator for diffusion A^d acting on such a function, $f = \sum_{M=0}^{K} 1_{\mathcal{X}(M)} f_M$, according to Eqn. 23 gives

$$\begin{aligned}\int_{\mathcal{X}} A^d f(x)\pi(x)dx \;=\; & \sum_{M=0}^{K} [\, -\int_{\mathcal{X}(M)} \frac{1}{2} < \nabla H_M(x), \nabla f_M(x) > \frac{e^{-H_M(x)}}{\mathcal{Z}} dx \\ & + \int_{\mathcal{X}(M)} \frac{1}{2} \left(\sum_{i=1}^{3n(M)} \frac{\partial^2 f_M(x)}{\partial p_i^2} \right) \frac{e^{-H_M(x))}}{\mathcal{Z}} dx\,].\end{aligned}$$

where $< \cdot, \cdot >$ stands for the vector dot-product and the gradients $\nabla H_M(x), \nabla f_M(x)$ are w.r.t the position (velocity) vector, an element of $R^{3n(M)}$. Integration by parts of the second term, with the fact that the functions f_M vanish at the boundary, results in a term which is negative of the first term. Therefore the given posterior π is the stationary density of the diffusion process.

We note that the curved nature of the torus requires the argument to be modified in sufficiently subtle ways. For details of such modifications to the manifolds associated with Lie groups see [30]. The jump part of the generator A^j is given by

$$A^j f(x) = q(x) \int_{\mathcal{X}(\mathcal{F}^1(x))} Q(x, dy)(f(y) - f(x)) \ ,$$

and computing the adjoint corresponding to the Eqn. 23 (see [1] for illustration) provdes the balance condition

$$q(x)\pi(x)dx = \int_{\mathcal{X}(\mathcal{F}^{-1}(x))} q(y, dx)\pi(y)dy \ , \tag{25}$$

where $\mathcal{F}^{-1}(x) \subset \aleph$ is the subset of models which can reach x in one jump transition. The jump parameters must satisfy Eqn. 25 for the density $\pi(x)$ to be stationary for the jump part of the process. For definiteness, assume $x \in \mathcal{X}(M)$ so that $x \equiv x(M)$ is an M-track configuration. Substituting for the transition measures from Eqns. 15 gives

$$\begin{aligned}
\pi(x(M))dx(M) \quad [\quad & \sum_{j=1}^{M} \int_{\mathcal{X}_0} q_s^b(x(M), x(M) \oplus_j y)dy + \sum_{j=1}^{M} q_s^d(x(M), \vartheta_{s(j)}x(M)) \\
+ \quad & \sum_{j=1}^{M+1} \int_{\mathcal{X}_0 \times \mathcal{A}} q_t^b(x(M), x(M) \oplus_j y(1))dy(1) + \sum_{j=1}^{M} q_t^d(x(M), \vartheta_{t(j)}x(M)) \\
+ \quad & \sum_{j=1}^{M} q_a(x(M), \vartheta_{a(j)}x(M)) \ ] \\
= dx(M) \ [& \sum_{j=1}^{M} \int_{\mathcal{X}_0} q_s^d(x(M) \oplus_j y, x(M))\pi(x(M) \oplus_j y)dy + \sum_{j=1}^{M} q_s^b(\vartheta_{s(j)}x(M), x(M))\pi(\vartheta_{s(j)}x(M))
\end{aligned}$$

$$+\sum_{j=1}^{M+1}\int_{\mathcal{X}_0\times\mathcal{A}} q_t^d(x(M)\oplus_j y(1),x(M))\pi(x(M)\oplus_j y(1))dy(1)+\sum_{j=1}^{M} q_t^b(\vartheta_{t(j)}x(M),x(M))\pi(\vartheta_{t(j)}x(M))$$
$$+\sum_{j=1}^{M} q_a(\vartheta_{a(j)}x(M),x(M))\pi(\vartheta_{a(j)}x(M))\ ]\ .$$

We will prove this equality treating only the first two summation terms from both sides, corresponding to the birth/death of track-segments; the treatment for the rest being similar. The jump moves considered here, birth/death of track-segments, are defined by Eqns. 20 and 21. Substituting the values for q_s^b, q_s^d from Eqn. 19,

L.H.S.:

$$\pi(x(M))\ \frac{dx(M)}{Z_S(1)}\ \frac{1}{5M}\cdot\sum_{j=1}^{M}\Big[\int_{\mathcal{X}_0} e^{-[L_M(x(M)\oplus_j y)-L_M(x(M))]_+}e^{-P_j(y|x(M))}dy$$
$$+e^{-[L_M(\vartheta_{s(j)}x(M))-L_M(x(M))]_+}\Big]$$
$$=\ \frac{dx(M)}{(\mathcal{Z})(Z_S(1))}\ \frac{1}{5(M)}\sum_{j=1}^{M}\Big[\int_{\Omega_>} e^{-L_M(x(M))}e^{-P_M(x(M))}e^{-P_j(y|x(M))}dy$$
$$+\int_{\Omega_\leq} e^{-L_M(x(M)\oplus_j y)}e^{-P_M(x(M))}e^{-P_j(y|x(M))}dy$$
$$+e^{-L_M(x(M))}e^{-P_M(x(M))}e^{-[L_M(\vartheta_{s(j)}x(M))-L_M(x(M))]_+}\Big]\ . \qquad (26)$$

R.H.S.:

$$\frac{dx(M)}{Z_S(1)}\ \frac{1}{5(M)}\sum_{j=1}^{M}\Big[\int_{\mathcal{X}_0} e^{-[L_M(x(M))-L_M(x(M)\oplus_j y)]_+}\pi(x(M)\oplus_j y)dy$$
$$+e^{-[L_M(x(M))-L_M(\vartheta_{s(j)}x(M))]_+}e^{-P_j(y|\vartheta_{s(j)}x(M))}\pi(\vartheta_{s(j)}x(M))\Big]$$
$$=\ \frac{dx(M)}{(\mathcal{Z})(Z_T(1))}\ \frac{1}{5M}\sum_{j=1}^{M}\Big[\int_{\Omega_>} e^{-L_M(x(M))}e^{-P_M(x(M)\oplus_j y)}dy$$
$$+\int_{\Omega_\leq} e^{-L_M(x(M)\oplus_j y)}\ e^{-P_M(x(M)\oplus_j y)}dy$$
$$+e^{-[L_M(x(M))-L_M(\vartheta_{s(j)}x(M))]_+}e^{-P_j(y|\vartheta_{s(j)}x(M))}e^{-L_M(\vartheta_{s(j)}x(M))}e^{-P_M(\vartheta_{t(j)}x(M))}\Big]\ , \qquad (27)$$

where

$$\Omega_{>} = \mathcal{X}_0 \bigcap \{y : L_M(x(M)) > L_M(x(M) \oplus_j y)\}$$

$$\Omega_{\leq} = \mathcal{X}_0 \bigcap \{y : L_M(x(M)) \leq L_M(x(M) \oplus_j y)\} \ .$$

Comparing Eqns. 26, 27 and combining the prior terms (i.e. $P_M(x(M) \oplus_j y) = P_M(x(M)) + P_j(y|x(M))$, and $P_M(x(M)) = P_M(\vartheta_{s(j)}x(M)) + P_j(y|\vartheta_{s(j)}x(M)))$, the equality is verified.

Q.E.D

References

[1] U. Grenander and M. I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society B*, 56(3):549–603, 1994.

[2] B. Gidas. Metropolis type monte-carlo simulation algorithms and simulated annealing. In *Trends of Contemporary Probability*, 1993. to appear.

[3] J. Besag and P. J. Green. Spatial statistics and bayesian computation. *J. Royal Statistical Society B*, 55:25–38, 1993.

[4] U. Grenander and M. I. Miller. Jump-diffusion processes for abduction and recognition of biological shapes. *Monograph of the Electronic Signals and Systems Research Laboratory*, 1991.

[5] Y. Bar-Shalom and E. Tse. Tracking in cluttered environment with probabilistic data association. *Automatica*, (11):451–460, 1975.

[6] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association.* Academic Press, 1988.

[7] Editor: Y. Bar-Shalom. *Multitarget-Multisensor Tracking.* Artech House, 1990.

[8] C. R. Rao, C. R. Sastry, and B. Zhou. Tracking the direction of arrival of multiple moving targets. *IEEE Transactions on Acoustics,Speech and Signal Processing*, accepted for publication, 1993.

[9] C. R. Sastry, E. W. Kamen, and M. Simaan. An efficient algorithm for tracking angles of arrival of moving targets. *IEEE Transactions on Acoustics,Speech and Signal Processing*, ASSP-39(No.1):242–246, 1991.

[10] C. K. Sword, M. Simaan, and E. W. Kamen. Multiple target angle tracking using using sensor array outputs. *IEEE Trans. AES.*, 26(2):367–373, 1990.

[11] W. Freiberger and U. Grenander. Computer generated image algebras. *Internation Federation Information Processing*, 68:1397–1404, 1969.

[12] U. Grenander. Advances in pattern theory: The 1985 Rietz lecture. *The Annals of Statistics*, 17:1–30, 1985.

[13] Nicholas Cutaia and Joseph A. O'Sullivan. Automatic target recognition algorithms using kinematic priors. *Electronic Signals and Systems Research Lab. Monograph*, 1994.

[14] Bernard Friedland. *Control System Design : An Introduction To State-Space Methods.* McGraw-Hill Book Company, 1986.

[15] K.V. Mardia. *Statistics of Directional Data.* Academic Press, London and New York, 1972.

[16] J. Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of Statistics*, 11:416–431, 1983.

[17] J. Rissanen. Stochastic complexity and modeling. *The Annals of Statistics*, 14, no.3:1080–1100, 1986.

[18] M.I. Miller and D. R. Fuhrmann. Maximum likelihood narrow-band direction finding and the EM algorithm. *IEEE Acoust. Speech and Signal Processing*, 38, No.9:560–577, 1990.

[19] A. Srivastava, M.I. Miller, and U. Grenander. Jump-diffusion processes for object tracking and direction finding. In *Proceedings of the 29th Annual Allerton Conference on Communication, Control and Computing*, pages 563–570, Urbana, Champaign, 1991. University of Illinois.

[20] A. Srivastava, N. Cutaia, M.I. Miller, J. A. O'Sullivan, and D. L. Snyder. Multi-target narrowband direction finding and tracking based on motion dynamics. In *Proceedings of the 30th Annual Allerton Conference on Communication, Control and Computing*, Urbana, Champaign, 1992. University of Illinois.

[21] M.I. Miller, R. S. Teichman, A. Srivastava, J.A. O'Sullivan, and D. L. Snyder. Jump-diffusion processes for automated tracking-target recognition. In *Proceedings of the Twenty-Seventh Annual Conference Conference on Information Sciences and Systems*, pages 617–622, Baltimore, Maryland, March 24-26 1993. Johns Hopkins University.

[22] R. Schmidt. *A signal subspace approach to multiple emitter location and spectral estimation.* Ph.D. Dissertation of Stanford University, Palo Alto, CA., Nov. 1981.

[23] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller. The use of maximum-likelihood estimation for forming images of diffuse radar-targets. In *Transactions of SPIE in Advanced Architectures and Algorithms*, San Diego, California, 1987.

[24] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller. The use of maximum-likelihood estimation for forming images of diffuse radar-targets from delay-doppler data. *IEEE Transactions on Information Theory*, 35(3):536–548, 1989.

[25] J.A. O'Sullivan, P. Moulin, and D.L. Snyder. Cramer-rao bounds for constrained spectrum estimation with application to a problem in radar imaging. In *Proceedings 26th Allerton Conference on Communication, Control, and Computing*, Champaigne, Urbana, October 1988. Urbana, IL.

[26] M.I. Miller, D.R. Fuhrmann, J.A. O'Sullivan, and D.L. Snyder. Maximum-likelihood methods for toeplitz covariance estimation and radar imaging. In Simon Haykin, editor, *Advances in Spectrum Estimation*, pages 145–172. Prentice-Hall, 1990.

[27] P. Moulin, J.A. O'Sullivan, and D.L. Snyder. A method of sieves for multiresolution spectrum estimation and radar imaging. *IEEE Transactions on Information Theory*, 1992.

[28] J.A. O'Sullivan, K. C. Du, R. S. Teichman, M.I. Miller, D.L. Snyder, and V.C. Vannicola. Radar target recognition using shape models. In *Proc. 30th Annual Allerton Conference on Communication,Control, and Computing*, pages 515–523, Urbana, Il., 1992. University of Illinois.

[29] I. I. Gihman and A. V. Skorohod. *Introduction to the Theory of Random Processes.* Saunders, Philadelphia, 1965.

[30] Y. Amit and M.I. Miller. Ergodic properties of jump-diffusion processes. *Monograph of the Electronic Signals and Systems Research Laboratory*, January 1993.

[31] A. Srivastava, M.I. Miller, and U. Grenander. Multi-target direction of arrival tracking. *IEEE Transactions on Signal Processing*, to appear, May 1995.

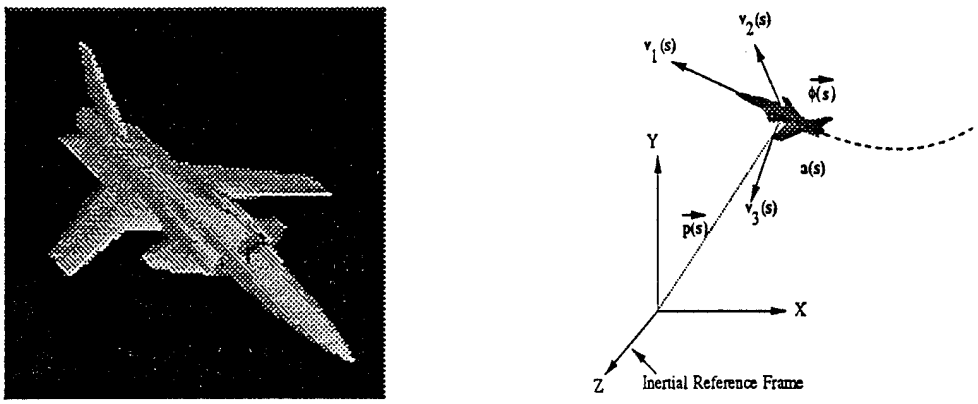


Figure 1: The left panel shows the 3-D target generator $g \in \mathcal{G}^o$ under a similarity transformations. The right panel shows the target located at position $p(s)$, oriented at $\phi(s)$ with velocities $v(s)$ resolved in the body frame coordinates.

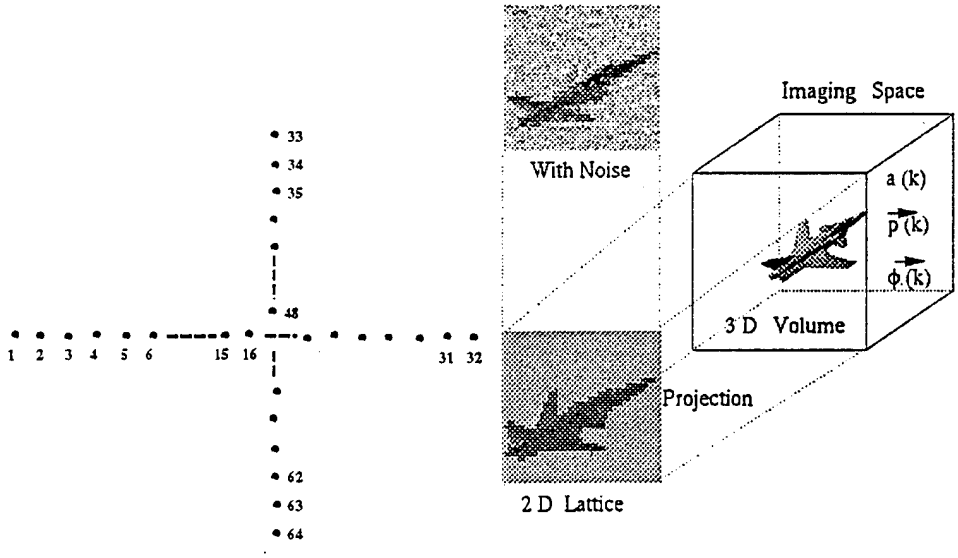


Figure 2: The left panel displays the cross array of isotropic sensors at half wavelength spacing, used to observe the angular location of the target. The right panel shows the far-field orthographic imaging system used for observing the targets at a high resolution.

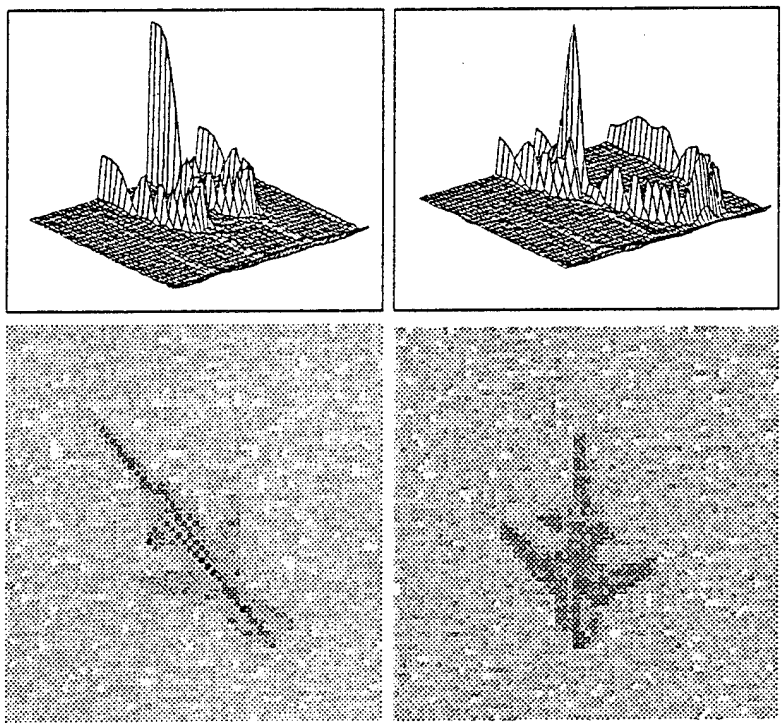


Figure 3: The upper panels show the azimuth-elevation power spectrum of the tracking data at two sample times. The lower panels display the high resolution data sets for the target at two different times during the flight path.

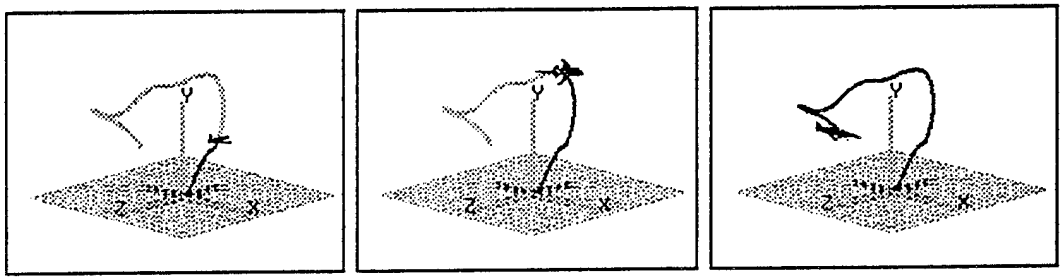


Figure 4: The actual track drawn in grey is observed by the ground based observation system. The track estimates are drawn overlapping in black at three stages of the algorithm.

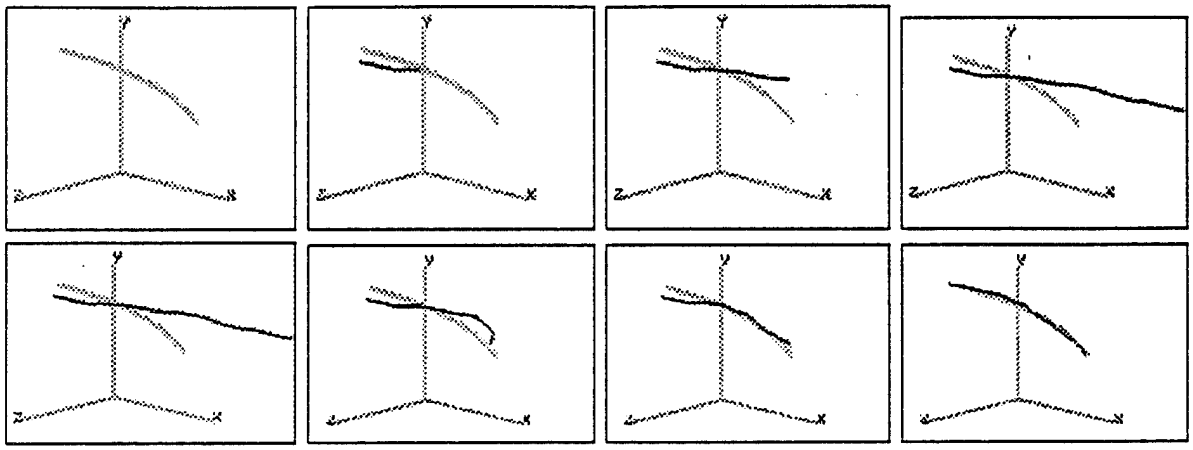


Figure 5: The upper panels show the sequence of jump moves adding segments to the estimated state from left to right, with the diffusion turned off. The lower panels show the continuous diffusion transformation aligning the estimated to the true track via the gradients on the posterior energy.

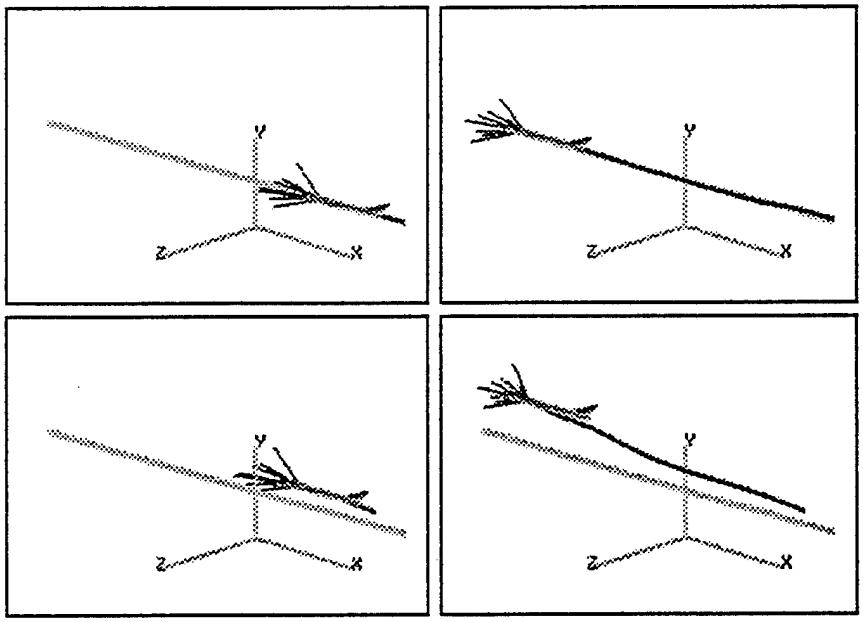


Figure 6: The four panels show candidates from the prior distribution for target path estimation with the high prior probability candidates forming a cone at track end for the algorithm to sample from. The upper panels show the prior with the diffusion on track parameters turned on; the lower panels have the diffusion turned off.

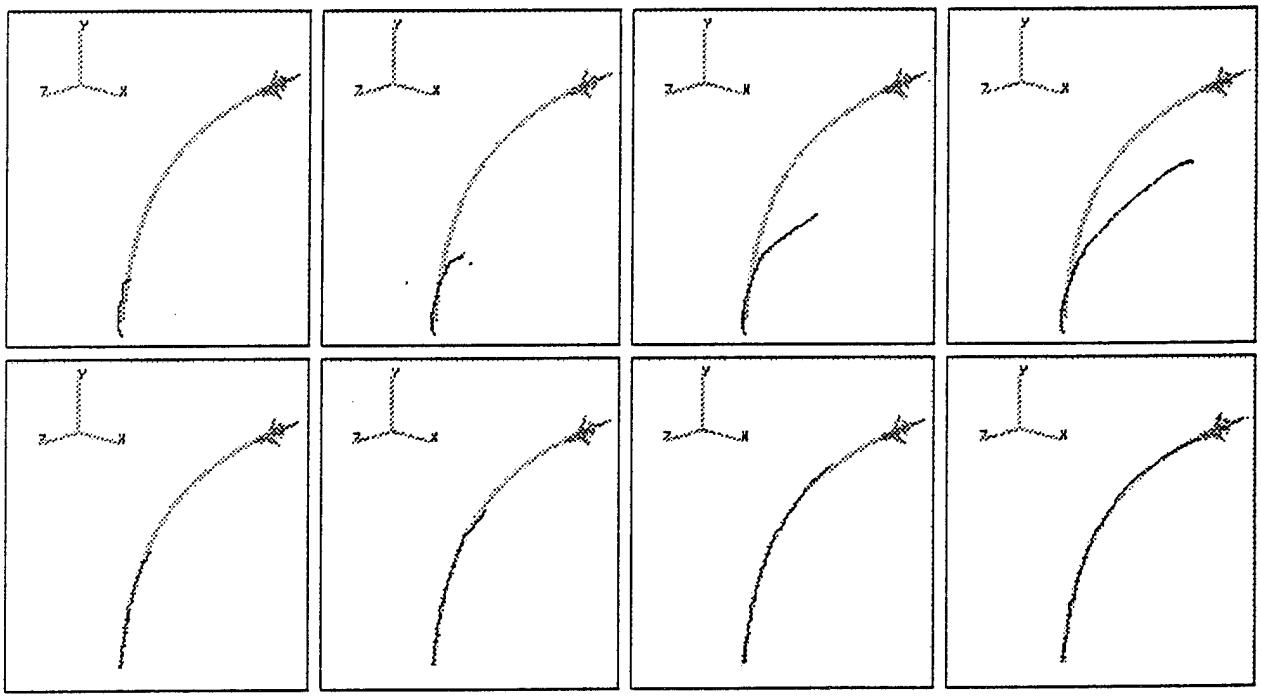


Figure 7: The upper panels display the estimated states at four times without the tracking prior information. The lower panels show the results of the dynamics based estimation algorithm with the prior included.

APPENDIX D

Multiple Target Direction of Arrival Tracking

A. Srivastava, M. I. Miller, and U. Grenander

Abstract—**A new algorithm is presented for estimating the directions of arrival of signals from an unknown number of moving signal sources. The parameter space is a countable union of Cartesian products of the torus, each product space corresponding to a different number of signals reaching the sensor array. A Bayesian posterior probability measure is defined on this parameter space by combining the sensor array manifold models with the von-Mises prior on source motion. The estimates are generated empirically using a random sampling algorithm based on *Jump-Diffusion* processes and the results are presented from an implementation on a DAP510 massively parallel computer.**

I. Introduction

This correspondence focuses on tracking the directions of arrival (DOA's) of signals emitted from multiple sources in remotely sensed, dynamically changing scenes. There are an unknown number of signal sources assumed moving in a 2-D plane containing a uniform linear array of passive, isotropic sensors. The goal is to track their angular locations relative to the array. Taking the minimum mean

Manuscript received June 4, 1994; revised October 14, 1994. This paper was presented in part at the 29th Annual Allerton Conference on Communication, Control, and Computing, 1991, University of Illinois at Urbana-Champaign, Illinois. This work was supported by the ARO DAAL03-92-G-0141, ONR N00014-94-1-0859, ONR N00014-92-J-1418, Rome Laboratory F30602-92-C-0004, ARO/MIT DAALO3-92-G-0115, ONR N00014-91-J-1021, ARL MDA972-93-1-0012, and NSF DMS-9217655.

A. Srivastava and M. I. Miller are with the Department of Electrical Engineering, Electronic Signals and Systems Research Laboratory, Washington University, St. Louis, MO 63130 USA.

U. Grenander is with the Division of Applied Mathematics, Brown University, Providence, Rhode Island USA.

IEEE Log Number 9410279.

squared error (MMSE) approach, we seek the conditional means of the statistics of targets' positions under the posterior density.

The posterior density being highly nonlinear in the track parameters precludes the closed-form analytic generation of conditional expectations. Therefore, we utilize recent advances in the use of random sampling methods for empirically generating estimates from complicated distributions [2], [3]. The random samples are generated by simulating a Markov process with the *ergodic property* that the empirical distribution of the samples converges to the posterior distribution as described in [1] and [4]. The algorithm searches through the connected parts of the parameter space (Cartesian products of the torus) with sample paths corresponding to the solutions of standard diffusion equations; across the disconnected parts of parameter space the jump process determines the transitions, satisfying *jump-diffusion dynamics* in such a way that the sample statistics converge to their expectation under the posterior. Such methods have been applied previously to the understanding of electron-microscope images containing sub-cellular structures such as mitochondrias and linear membranes [1].

There exists a substantial literature on estimating the DOA's of moving/stationary sources recorded by sensor arrays [5], [6] with the nonlinear problem solved using gradient-based techniques and eigenvalue analyses in mostly maximum-likelihood settings. Most of this work assumes knowledge of the number of targets which, in general, are time varying and unknown *a priori*. The jump-diffusion based sampling algorithm jointly estimates the location parameters along with the track number and track lengths.

II. Parameter Space and Bayesian Posterior

Our interest lies in the angular locations (azimuth or elevation) $x \in \mathcal{T}(1) \equiv [0, 2\pi], 0, 2\pi$ identified, of the sources assumed to be present in a plane (2-D space) containing a narrowband uniform linear array (Fig. 1) of P sensors, with known array manifold as described in [7]. Notice the problem setup inherently involves range ambiguity. The sequence of (angular) locations a target attains during its motion forms a track, the individual locations at discrete sample times form track-segments, and the number of segments constituting a sampled track is its track-length. The parameter vector $\vec{x}_t^{(m)}$ of the mth target over the observation period $[t_0, t]$ is an element of $(\mathcal{T}(1) \cup \{\sharp\})^{[t_0,t]}$, where $\{\sharp\}$ stands for the target's absence in the observation space. The M-track parameter vector $\vec{x}_t(M)$ becomes an element of the space $\mathcal{X}_t(M) \equiv (\mathcal{T}(1) \cup \{\sharp\})^{M[t_0,t]}$. Since M is unknown *a priori*, the complete space is defined to be the countable union $\mathcal{X}_t = \cup_{M=0}^{\infty} \mathcal{X}_t(M)$. For later convenience, define $x^{(m)}(\tau), \tau \in [t_0, t]$ as the location of the mth target at time τ and $n(M) \in \aleph$ as the total number of location parameters (or segments) describing an M-track scene.

Under the array manifold data model, the relative phase lags of the signals arriving at the different sensor elements are known functions of the source locations and sensor geometry. For a uniform linear array of sensors at half-wavelength spacing the measurement vector, for $\tau \in [t_0, t]$, is given by

$$\vec{y}(\tau) = \sum_{m=1}^{M} \vec{d}(x^{(m)}(\tau)) 1_{\mathcal{T}(1)}(x^{(m)}(\tau)) s^{(m)}(\tau) + \vec{n}(\tau) \quad (1)$$

where $s^{(m)}(\tau)$ is the deterministic signal amplitude of the mth source at time τ, $\vec{n}(\tau)$ is a $P \times 1$ complex Gaussian noise vector of Goodman class, $CN(0, \sigma^2)$, and $\vec{d}(x^{(m)}(\tau))$ is the direction vector

$$\vec{d}(x) = [1 e^{-i\pi\cos(x)}, \cdots, e^{-i(P-1)\pi\cos(x)}]^T, i = \sqrt{-1}.$$

The indicator function $1_{\mathcal{T}(1)}(x^{(m)}(\tau))$ is defined to be one if $x^{(m)}(\tau) \in \mathcal{T}(1)$ and zero if $x^{(m)}(\tau) = \sharp$. It selects the targets

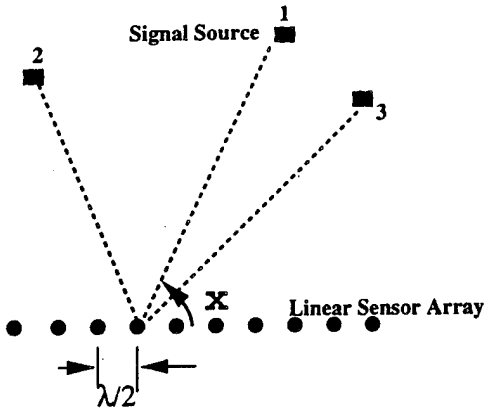


Fig. 1. A uniform linear array of isotropic sensors at half wavelength spacing used to observe the angular locations of various sources.

that contribute to the array manifold at time τ. Assume the targets are stationary during some fundamental data sampling intervals, with their locations given by the values on some index set $\{\tau_j\}_{j=1,\cdots,J}, J$ being the total number of sample points in the observation interval $[t_0, t]$. Then, for an M-track configuration $\vec{x}_t(M) \in \mathcal{X}_t$, the Gibb's energy term for the data likelihood becomes

$$L_t(\vec{x}_t(M)) = \frac{1}{\sigma^2} \sum_{j=1}^{J} \left| \vec{y}(\tau_j) - \sum_{m=1}^{M} \vec{d}(x^{(m)}(\tau_j)) 1_{\mathcal{T}(1)}(x^{(m)}(\tau_j)) s^{(m)}(\tau_j) \right|^2.$$

The prior measure is formed using the von-Mises density on the torus $x \in \mathcal{T}(1)$ [8], $F(x) = (1/2\pi I_0(\kappa)) e^{\kappa\cos(x-\bar{x})}$, where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero, $\kappa > 0$ is the concentration parameter, and $\bar{x}$ is the mean of x. The process is made Markov by assigning the previous state as the mean of the current state. In an M-track scene, the smoother tracks are favored according to the energy function

$$P(\vec{x}_t(M)) = \kappa \sum_{j=2}^{J} \sum_{m=1}^{M} \cos(x^{(m)}(\tau_j) - x^{(m)}(\tau_{j-1})).$$

To encourage simple model deductions, a complexity term similar to Rissanen's [9], [10] can be an effective prior for integer-valued random variables such as the track lengths and the source numbers.

The posterior distribution μ_t is then of the Gibb's type supported on $\mathcal{X}_t$, i.e., for all sets $\mathcal{A} \subset \mathcal{X}_t$, Lebesgue measurable

$$\mu_t(\mathcal{A}) = \sum_{M=0}^{\infty} \mu_t(\mathcal{A} \cap \mathcal{X}_t(M)) = \sum_{M=0}^{\infty} \int_{\mathcal{A} \cap \mathcal{X}_t(M)} \frac{e^{-H_t(x_t(M))}}{Z_t} dx_t(M)$$

with $H_t(x_t(M)) = P_t(\vec{x}_t(M)) + L_t(\vec{x}_t(M)) - \log(p_M)$, $\Sigma_{M=0}^{\infty} p_M = 1$, and $Z_t = \Sigma_{M=0}^{\infty} \int_{\mathcal{X}_t(M)} e^{-H_t(x_t(M))} dx_t(M)$.

III. Conditional Mean Estimation

We construct a Markov process $X(s)$ for sampling from the posterior to obtain the conditional mean estimates empirically. The diffusion component of $X(s)$ continuously winds through the parameter subspaces $\mathcal{X}_t(M), M = 0, 1, 2, \cdots$, following the stochastic gradients of the posterior potential $H_t(\cdot)$. The jump component deduces the target numbers and the track lengths via a family of discontinuous transformations on the scene, corresponding to i) estimating source or track numbers by hypothesizing the birth/death of the tracks, or ii) changing the track lengths by hypothesizing the

birth/death of the track-segments. It is the fundamental difference between diffusions (almost surely continuous sample paths) and jump processes (large moves in parameter space in small time) that allows us to explore the very different connected and disconnected nature of the parameter space $\mathcal{X}_t$. The following analysis is carried out for the fixed observation period $[t_0, t]$ so the subscript t is suppressed without ambiguity.

The **jump process** includes a variety of jump moves, through the disconnected subspaces, denoted by the set of operators

$$\mathcal{F} = \{\vartheta^d_{t(m)}, \vartheta^d_{s(m)}, \vartheta^b_{t(m)}, \vartheta^b_{s(m)}\}$$

according to

$$\begin{aligned}
\vartheta^d_{t(m)}&: \mathcal{T}(1)^{n(M)} \to \mathcal{T}(1)^{n(M)-1},\\
\vartheta^d_{s(m)}&: \mathcal{T}(1)^{n(M)} \to \mathcal{T}(1)^{n(M)-1}\\
\vartheta^b_{t(m)}&: \mathcal{T}(1)^{n(M)} \to \mathcal{T}(1)^{n(M)+1},\\
\vartheta^b_{s(m)}&: \mathcal{T}(1)^{n(M)} \to \mathcal{T}(1)^{n(M)+1}.
\end{aligned}$$

The first two are the deletion operators removing the mth track of length one and the last segment from mth track, respectively. The last two are the birth operators adding a track (indexed by m) to the scene and a segment at the end of the mth track, respectively. Notice in one jump move, the addition/deletion of only unit length tracks or single track segments is allowed.

The jump transitions follow the probability measure $Q(\vec{x}, d\vec{y}) = q(\vec{x}, d\vec{y})/\int_{\mathcal{X}} Q(\vec{x}, d\vec{y})$ with the standard definition (e.g. [11])

$$q(\vec{x}, d\vec{y}) = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(\Pr\{X(s+\epsilon) \in dy | X(s) = \vec{x}\} - 1_{d\vec{y}}(\vec{x})).$$

Let $\mathcal{F}^1(\vec{x})$ be the set of models that can be reached from $\vec{x}$ in one jump move, and $\mathcal{X}(\mathcal{F}^1(\vec{x}))$ the space containing the configurations of these types. The intensity of the jump process at $\vec{x}$ is given by $q(\vec{x}) = \int_{\mathcal{X}(\mathcal{F}^1(\vec{x}))} q(\vec{x}, d\vec{y})$. The transition measures are made singular with respect to the Lebesgue measures on the respective subspaces which the jump transformations move into. For this, the part of the state that is not being added or deleted remains unchanged. The resulting measures on $\mathcal{X}(\mathcal{F}^1(\vec{x}(M)))$ are

$$\begin{aligned}
&q(\vec{x}(M), d\vec{y}(M+1))\\
&\quad = \sum_{m=1}^{M+1} q^b_t(\vec{x}(M), \vec{y}(M+1))\delta_{\vec{x}(M)}(d(\vartheta^d_{t(m)}\\
&\qquad \cdot \vec{y}(M+1)))\, d\vec{y}(1),\\
&q(\vec{x}(M), d\vec{y}(M))\\
&\quad = \sum_{m=1}^{M} q^b_s(\vec{x}(M), \vec{y}(M))\delta_{\vec{x}(M)}(d(\vartheta^d_{s(m)}\vec{y}(M)))\, dy\\
&\qquad + \sum_{m=1}^{M} q^d_s(\vec{x}(M), \vec{y}(M))\delta_{\vartheta^d_{s(m)}\vec{x}(M)}(d\vec{y}(M)),\\
&q(\vec{x}(M), d\vec{y}(M-1))\\
&\quad = \sum_{m=1}^{M} q^d_t(\vec{x}(M), \vec{y}(M-1))\delta_{\vartheta^d_{t(m)}\vec{x}(M)}(d\vec{y}(M-1))
\end{aligned}$$

where $q^b_t, q^d_t, q^b_s, q^d_s$ are the intensities associated with the transformations $\vartheta^b_{t(m)}, \vartheta^d_{t(m)}, \vartheta^b_{s(m)}, \vartheta^d_{s(m)}$, respectively. We choose the jump intensities in such a way (via the algorithm described below) that the Bayesian posterior satisfies the backward Kolmogoroff condition for stationary measure [1] given by $q(x)\mu(dx) = \int_{\mathcal{X}} q(y)Q(y, dx)\mu(dy)$.

To analyze the **diffusion process**, we utilize results from [12], [1], and [4] to construct stochastic flows on Lie-manifolds such as the torus. For the subspace associated with M tracks having $n(M)$ segments the diffusion flows through the manifold $\mathcal{X}(M) \equiv \mathcal{T}(1)^{n(M)}$. It is essentially a stochastic gradient, in each of the subspaces $\mathcal{X}(M)$, on the posterior potential $H(\vec{x}(M))$ generated by a Langevin's stochastic differential equation (SDE)

$$X(s) = \left[X(0) + \int_0^s -\tfrac{1}{2}\nabla H(X(\tau))\, d\tau + W(s)\right]_{\text{mod } 2\pi} \tag{2}$$

where $[\cdot]_{\text{mod } 2\pi}$ is taken componentwise, $W(s) \in \Re^{n(M)}$ is the standard vector Wiener processes, and $\nabla H(X(\tau)) \in \Re^{n(M)}$ is the gradient with respect to $\vec{x}(M)$. The resulting backward Kolmogoroff operator A for the Markov process we have constructed satisfies the condition of stationarity given by $\int_{\mathcal{X}} Af(x)\mu(dx) = 0$ for all $f(\cdot) \in \text{domain}(A)$, as shown in [1], [4].

With this choice of jump and diffusion parameters we have also proven, in [1], [4], and [13], that $X(s)$ converges in variation norm to the posterior measure $\mu(\cdot)$, implying that the empirical averages of functions on the sample paths converge to their expectations under the posterior density, i.e., $1/n\,\Sigma_{\tau=1}^n f(X(\xi_\tau)) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} f(x)\mu(dx)$.

IV. Algorithm and Simulation Results

The simulation time ξ is different from the data sample times $\tau \in \{\tau_j\}_{j=1}^J$, which form a tiling of $[t_0, t]$. The jump-diffusion process is constructed as follows. Associate the probabilities $(1/4)/K, (1/4)/K, (1/4K_T)/K, (1/4K_S)/K$ with the set of jump moves t^d, s^d, t^b, s^b, respectively, where $K_S = 2\pi I_0(\kappa)$ is the von-Mises normalizer, $K_T = 2\pi^J$ is the normalizer for uniform prior on the set $\mathcal{T}(1)^J$ of possible track seeds, and

$$K = 1/4 + 1/4 + 1/4K_T + 1/4K_S.$$

Initialize with $\xi_0 = 0, i = 0$.

1) Generate an exponential random variable u with mean one. Let $X(\xi_i) \in \mathcal{X}(M)$ for some M. For $\xi \in [\xi_i, \xi_i + u), X(\xi) \in \mathcal{X}(M)$ follows the SDE (2).
2) At $\xi_{i+1} = \xi_i + u$, define $\vec{x} = X(\xi^-_{i+1})$, draw one of the 4 possible jump choices from the set $\{t^b, s^b, t^d, s^d\}$, with probabilities defined above, and choose one of the following.
   - If $t^{(b)}$, draw a unit length track (track seed) from a uniform prior $1/2\pi^J$, draw $m \in \{1, \cdots, M+1\}$ uniformly, $\vec{x}_{\text{new}} = \vartheta^b_{t(m)}\vec{x}, M_{\text{new}} = M+1$.
   - Else if $t^{(d)}$, draw $m \in \{1, \cdots, M\}$ uniformly, if mth track is unit length, then $\vec{x}_{\text{new}} = \vartheta^d_{t(m)}\vec{x}, M_{\text{new}} = M-1$.
   - Else if $s^{(b)}$, draw a segment from prior and draw $m \in \{1, \cdots, M\}$ uniformly: $\vec{x}_{\text{new}} = \vartheta^b_{s(m)}\vec{x}, M_{\text{new}} = M$.
   - Else if $s^{(d)}$, draw $m \in \{1, 2, \cdots, M\}$ uniformly, $\vec{x}_{\text{new}} \leftarrow \vartheta_{s(m)}\vec{x}, M_{\text{new}} = M$.
3) If $[L_M(\vec{x})) - L_{M_{\text{new}}}(\vec{x}_{\text{new}})] > 0, X(\xi_{i+1}) \leftarrow x_{\text{new}}$; else $X(\xi_{i+1}) \leftarrow \vec{x}_{\text{new}}$ with probability $e^{-[L_{M_{\text{new}}}(\vec{x}_{\text{new}}) - L_M(\vec{x})]}$, and, $X(\xi_{i+1}) \leftarrow \vec{x}$ with probability $1 - e^{-[L_{M_{\text{new}}}(\vec{x}_{\text{new}}) - L_M(\vec{x})]}$.
4) $i \leftarrow i+1$, return to 1.

The algorithm has been implemented on a massively parallel AMT-DAP 510, a SIMD machine with a 32 × 32 mesh of processors. The targets were assumed present at all times, their number not known *a priori*. Also, 20 snapshots are taken for each target location with number of sensors $P = 10$ at S/N of 2dB. Fig. 2 shows the samples generated by the jump-diffusion algorithm at

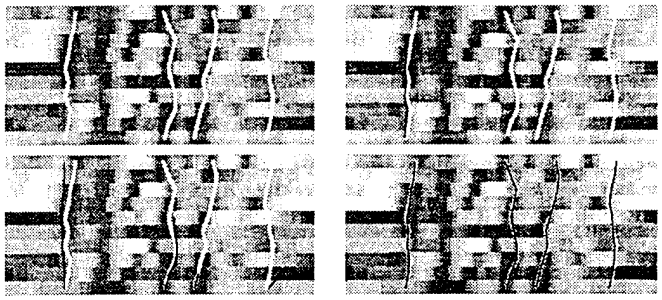


Fig. 2. Estimation of a four-track scene: The upper left panel shows the true tracks in white with the spatial power spectrum generated via MVDR from the noisy data. The other three panels show the successive stages of the algorithm (estimates drawn in black) with the final result shown in the lower right panel.

four successive stages for a four-track configuration, deducing both the target paths as well as the number of targets. The vertical axis represents real time progression of the sources moving across the array, while the horizontal axis parameterizes the angular location of targets in $\mathcal{T}(1)$. The spatial power spectrum as measured using the *minimum variance distortionless response* (MVDR) beamformer [14] is shown in the background, with darkness representing higher spatial power. Superimposed are the actual tracks in white with the estimated tracks shown in black. Notice the disparity between the true track and the estimates following the jumps at the initial stages, with the state finally brought into alignment by the diffusion transformation.

In situations where the tracks intersect or the sources are in close vicinity, it is the prior on track formation that contributes to the correct track assignment. The likelihood being the superposition of components from various sources doesn't discriminate between sources but a prior deriving information from the track histories and motion analysis does. One such prior, based on Newtonian equations of motion, is presented in [15] where the 3-D tracking of airplanes is analyzed.

REFERENCES

[1] U. Grenander and M. I. Miller, "Representations of knowledge in complex systems," *J. Royal Statistical Soc.*, vol. 56, no. 3, 1994.

[2] B. Gidas, "Metropolis type Monte-Carlo simulation algorithms and simulated annealing," in *Topics in Contemporary Probability and its Applications*, J. Laurie Snell, Ed. CRC Pr, 1995.

[3] J. Besag and P. J. Green, "Spatial statistics and Bayesian computation," *J. Royal Statistical Soc. B*, vol. 5, pp. 25–38, 1993.

[4] Y. Amit, and M. I. Miller, "Ergodic properties of jump-diffusion processes," *Monograph Electron. Signals Syst. Res. Laboratory*, Washington Univ., St. Louis, MO, 1992.

[5] C. R. Rao, C. R. Sastry, and B. Zhou, "Tracking the direction of arrival of multiple moving targets," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 42, no. 5, May 1994.

[6] C. K. Sword, M. Simaan, and E. W. Kamen, "Multiple target angle tracking using sensor array outputs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, no. 2, pp. 367–373, 1990.

[7] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, Nov. 1981.

[8] K. V. Mardia, *Statistics of Directional Data*. New York: Academic Pr, 1972.

[9] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Annals Statistics*, vol. 11, no. 2, pp. 416–431, 1983.

[10] ——, "Stochastic complexity and modeling," *Annals Statistics*, vol. 14, no. 3, pp. 1080–1100, 1986.

[11] I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes*. Philadelphia, PA: Saunders, 1969.

[12] Y. Amit, "A multiflow approximation to diffusions," *Stochastic Processes and their Applications*, vol. 37, no. 2, pp. 213–238, 1991.

[13] A. Srivastava, "Automated target tracking and recognition using jump diffusion processes," M.S. thesis, Washington Univ., St. Louis, MO, Dec. 1993.

[14] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," in *Proc. IEEE*, vol. 57, 1969, p. 1408.

[15] M. I. Miller, A. Srivastava, and U. Grenander, "Conditional-expectation estimation via jump-diffusion processes in multiple target tracking/recognition," *IEEE Trans. Signal Processing*, accepted for publication, Nov. 1994.

Representations of Knowledge in Complex Systems

BY

ULF GRENANDER and MICHAEL I. MILLER

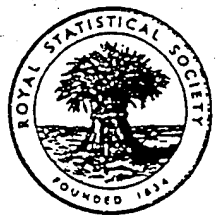
Reprinted from

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY

SERIES B (METHODOLOGICAL)

Volume 56, No. 4, 1994

(pp. 549–603)



Printed for Private Circulation

1994

J. R. Statist. Soc. B (1994)
56, *No.* 4, *pp.* 549–603

Representations of Knowledge in Complex Systems

By ULF GRENANDER and MICHAEL I. MILLER†

Brown University, Providence, USA *Washington University, St Louis, USA*

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 20th, 1993,* Professor V. S. Isham *in the Chair*]

SUMMARY

Modern sensor technologies, especially in biomedicine, produce increasingly detailed and informative image ensembles, many extremely complex. It will be argued that pattern theory can supply mathematical representations of subject-matter knowledge that can be used as a basis for algorithmic 'understanding' of such pictures. After a brief survey of the basic principles of pattern theory we shall illustrate them by an application to a concrete situation: high magnification (greater than 15000×) electron micrographs of cardiac muscle cells. The aim is to build algorithms for automatic hypothesis formation concerning the number, location, orientation and shape of mitochondria and membranes. For this we construct a pattern theoretic model in the form of a prior probability measure on the space of configurations describing these hypotheses. This measure is synthesized by solving sequentially a jump–diffusion equation of generalized Langevin form. The jumps occur for the creation–annihilation of hypotheses, corresponding to a jump from one continuum to another in configuration (hypothesis) space. These continua (subhypotheses) are expressed in terms of products of low dimensional Lie groups acting on the generators of a template. We use a modified Bayes approach to obtain the hypothesis formation, also organized by solving a generalized Langevin equation. To justify this it is shown that the resulting jump–diffusion process is ergodic so that the solution converges to the desired probability measure. To speed up the convergence we reduce the computation of the drift term in the stochastic differential equation analytically to a curvilinear integral, with the random term computed almost instantaneously. The algorithms thus obtained are implemented, both for mitochondria and membranes, on a 4000 processor parallel machine. Photographs of the graphics illustrate how automatic hypothesis formation is achieved. This approach is applied to deformable neuroanatomical atlases and tracking recognition from narrow band and high resolution sensor arrays.

Keywords: JUMP-DIFFUSION RANDOM SAMPLING; PATTERN THEORY; SHAPE RECOGNITION

1. INFERENCE IN COMPLEX SYSTEMS

The object of statistics is information. The objective of statistics is the understanding of information contained in data. To achieve such understanding statistics employs a variety of methods, one of the most powerful being the graphical display of characterizing functions derived from the data. We can speculate that a reason for this is that the visual processing in man is so formidable, not only in terms of its computing power but especially in its ability to organize its inputs into coherent structures. How such a logical organization is carried out is largely unknown but it seems plausible that it relies heavily on access to memory of enormous size compared with current computer memories: and not just memory, in some sense smart

† *Address for correspondence*: Department of Electrical Engineering, Washington University, St Louis, MO 63130, USA.
E-mail: mim@hyperion.wustl.edu

 0035-9246/94/56549

(active) memory. Whether or not this is so, it has served as a guiding principle in our work.

In an emerging field, supported by little or no theory, the choice of characterizing functions to be graphically displayed has to be done in an exploratory manner, aided by intuition and by informal guesses. As the field matures and theory develops, the researcher is aided in selecting functions, the functions which point to characteristic features strengthening or modifying the theory. At a still later stage the theory may be mathematically codified leading to automated procedures for making the inductions from data.

As an example take statistical signal processing. Of course this discipline can be traced back far, to Helmholtz and earlier, but here we are thinking of its history in this century. The advent of the cathode ray tube and the oscilloscope must have been invaluable tools for visualizing signals to understand their structure. Neurophysiological findings do not contradict the assertion that we do some sort of Fourier analysis while watching a waveform on the screen. The development of communication engineering from the 1920s onwards consisted in part of formalizing the observed, more or less noisy, signals utilizing ideas from Fourier analysis, stationary stochastic processes, Toeplitz forms, Bayesian inference and statistical mechanics to mention a few. Eventually this resulted in virtually automated procedures for the detection and understanding of noisy signals.

Signal processing is one of the great success stories of statistics. It is natural to ask why. We believe that it was because the pioneers in the field managed to construct representations of signal ensembles, models that were realistic and at the same time tractable both analytically and computationally (by analogue devices). Today these models are familiar, they look simple and natural, but in a historical perspective the phenomena must have appeared highly complex and bewildering.

But what are today's challenges in signal, data and pattern analysis? The last decades have witnessed a revolutionary development in the construction of sensors for capturing pictures: computerized tomography (CT), magnetic resonance (MR) imaging, optical sectioning, laser radars (range finders) and others. This has enabled researchers in the natural sciences, in particular biology and medicine, to acquire detailed images carrying vast amounts of information. The diagnostician, a user of these remarkable technologies, inspects the pictures looking for structure, perhaps watching out for abnormalities or unexpected behaviour. They rely on their training in anatomy, histology, cytology, etc. to understand what they see. One complication is, for certain modalities, that the pictures are quite noisy. Another is that they give only indirect information. For example, in CT the raw data sinogram is almost incomprehensible without computer processing. Or, the data may consist of two-dimensional slices whereas the anatomy is three dimensional, and so on.

The main difficulty we believe is that the anatomies (or other structures) form highly complex systems. Browsing through an anatomical text-book, e.g. Netter's (1980) beautiful atlas, one is overwhelmed by the awesome amount of information. Say that we have the ambition of creating algorithmic tools which help the diagnosticians by carrying out some of the time consuming labour while leaving the final decision to their judgment. To arrive at more than *ad hoc* algorithms the subject-matter knowledge must be expressed precisely and as compactly as possible. How can such empirical knowledge be represented in mathematical form, including both structure and the all-important variability?

This task is orders of magnitude bigger than modelling signal ensembles. For the latter a few parameters are needed, perhaps means and variances for Gaussian noise, or the spectral density of a signal source, and so on. The picture ensembles that we are now concerned with are so complex that megabytes of constants must be used to represent typical structure and more to describe variability, at least for the more challenging situations. These ensembles have little structure—they are the result of evolutionary incidents. Biology is not physics.

Here pictorial patterns (which are no accident) and biomedical images have been emphasized. Or putting the question more generally: how can *knowledge in complex systems be represented*, and how can such representations be studied analytically and exploited to give algorithmic solutions?

General pattern theory, a discipline that was initiated in the late 1960s, is intended to provide answers to this question. It gives an algebraic framework (*image algebras*) for describing patterns as structures regulated by rules, both local and global. Probability measures are sometimes superimposed on the image algebras to account for variability of the patterns. The resulting *regular structures* serve as the mathematical basis from which inference algorithms are *derived from first principles.*

Pattern theory borrows ideas and methods from algebra, probability, statistics and analysis, and is related to image processing, computer vision and pattern recognition. It differs from pattern recognition in that the latter emphasizes the construction of recognition algorithms whereas in pattern theory the representation of subject-matter knowledge is the centre of attention.

1.1. *Specific Example: Subcellular Shapes in Electron Micrographs*

To focus discussion, Fig. 1 shows electron micrograph sections of ventricular cardiac myocyte cells at 15000× magnification (Miller *et al.*, 1985). Each panel exhibits ultrastructural variability common to all myocyte cells which are composed predominantly of the 'energy producing' mitochondria (dark closed structures). There are clearly three kinds of variability which are common to biological specimens. First is the *structural or shape variability* associated with shape and scale variation. The second is the *internal constituent variability* of the textures making up the shapes. The third is the *complexity variability* due to the varying numbers of shapes, the number not known *a priori.* The challenge for the pattern theoretic approach is the construction of models which represent these variabilities in a mathematically precise way.

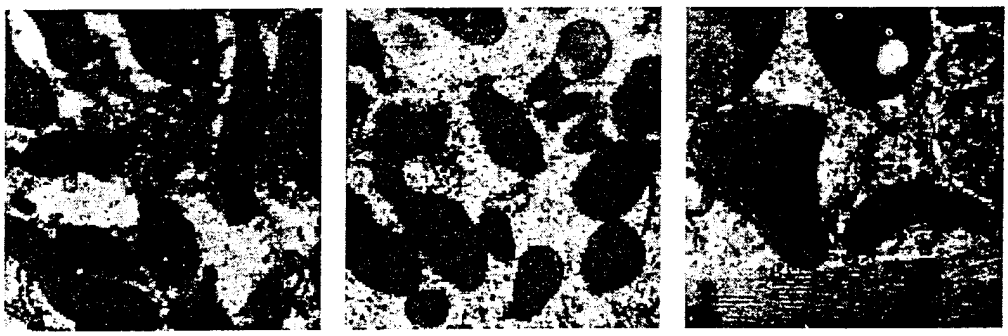


Fig. 1. Three electron microscope images at 15 000× magnification

The metric pattern theory accommodates the shape variability by defining *organelle templates* corresponding to fixed graphs with *rigid regularity* forcing global connectedness of the boundary. The templates are made flexible by attributing the templates with geometric group operators—scale, rotation and translation—resulting in rubber-band-like transformations of the rigid templates. The textural variability and sensor noise are accounted for via likelihoods constructed from Markov random field (MRF) models of the textured interiors. The complexity variability of arbitrary numbers of shapes is accommodated via the construction of a sample space of *scenes*. A scene consists of numbers of template shapes with their associated group transformations. Choosing a scene corresponds to choosing both the parameters determining the shapes as well as the *model order* (Rissanen, 1987) most representative of the data.

1.2. *Sequential Inference via Random Sampling*

The method of inference is Bayesian based starting from a posterior density relating the scenes to the observed sensor data. The inference method is a random process following *jump–diffusion dynamics* which draws samples from the posterior, thereby allowing for the empirical generation of conditional expectations of the various unknowns: means, covariances, etc. Since the model order is unknown, the posterior is defined over a countable union of real spaces, each space a variable parameterization of the pattern theoretic model. The motivation for introducing random sampling algorithms following jump–diffusion dynamics is to accommodate the very different continuous and discrete components of the object discovery process. Given a fixed number of objects in a scene the algorithm reshapes the templates to allow for the local variability of each shape in the data. The reshaping is performed continuously by using Langevin stochastic differential equations (SDEs), in which the state vector stochastically follows gradients of the posterior over shape space. The second part of the inference samples scenes of varying object number by using discontinuous jump moves which add objects, remove objects and fuse and split objects. The jump intensities are determined by the posterior density, with the diffusion equation governing the dynamics between jumps. A significant result of this work is that by appropriate choice of the jump–diffusion transition dynamics the resulting non-homogeneous Markov random sampler converges to the posterior implying that empirical averages converge to their expectations.

1.3. *Paper Lay-out*

With this introduction the main concepts of pattern theory are presented in Section 2. Then in Section 3 theorems on convergence of the random sampling algorithm are proven, with Section 4 focusing on the application to biological shapes in electron micrographs. Section 5 gives both computational details on the single-instruction, multiple-data (SIMD) implementation as well as experimental results, with Section 6 concluding with applications to deformable neuroanatomical atlases and tracking and recognition on passive sensor arrays.

2. PATTERN THEORETIC BACKGROUND

Global regularity is introduced into the representations (regular structures) via a family Σ, the connection type, of finite graphs, sometimes directed, sometimes not.

Let σ, a *connector*, be the symbol for a graph in Σ and $n = n(\sigma)$ be the number of sites in σ.

Often σ is fixed and possibly also known beforehand but this is not always the case; that will complicate the inferences as will be exemplified below. At each site, a mathematical object is placed, a *generator*, from some set $\mathcal{G}$, the *generator space*. The generators appear in many forms: they can mean geometric objects such as vectors and surface elements, or rewriting rules in language theory or computational modules in studies of algorithms. Whatever they are we insist that they carry *bonds*, information used to communicate to their neighbours on σ, and establish local regularity rules.

For a given generator g denote its bonds by $\beta_1, \beta_2, \ldots \beta_\omega$, where $\omega = \omega(g)$ is in algebraic parlance the *arity* of g. The β-values come from some bond value space $\mathcal{B}$, again of fairly general nature that will be specialized for each application.

The *local regularity* then takes the following form. Consider two generators g' and g'' situated at neighbouring sites in the connector σ so that a graph segment connects a bond $\beta_{j'}$ of g' with a bond $\beta_{j''}$ of g''. For each such segment, where two bonds meet, it is required that a *bond relation* $\rho(\ ,\ )$

$$\rho: \mathcal{B} \times \mathcal{B} \to \{\text{TRUE, FALSE}\} \tag{1}$$

takes the value $\rho(\beta_{j'}, \beta_{j''}) = \text{TRUE}$. When this is true everywhere in a *configuration*

$$c = \sigma(g(1), g(2), \ldots, g(n)) \tag{2}$$

then c is said to be regular. The above formalism is intended to bring out the algebraic structure of the resulting *configuration space* $\mathcal{C}(\mathcal{R})$ over the regularity $\mathcal{R} = \langle \mathcal{G}, \Sigma, \rho \rangle$.

Typically patterns allow natural invariances. They will be formalized through a *similarity group* $\mathcal{S}$ of bijective transformations acting on the generator space

$$s: \mathcal{G} \leftrightarrow \mathcal{G}, \qquad s \in \mathcal{S} \tag{3}$$

Often several similarity groups will be used on the same configuration space. A similarity s can be naturally extended to $\mathcal{C}(\mathcal{R})$ which receives an algebraic structure (see Grenander (1981), chapter 3).

The configurations are mathematical abstractions, typically not observable even in principle, with their relation to observables captured by some sensor technology and expressed by an *identification rule R*. Such a rule, assumed to be an equivalence relation, partitions $\mathcal{C}(\mathcal{R})$ into equivalence classes denoted I, the (pure) images, together forming an *image algebra* $\mathcal{I}$, $I \in \mathcal{I}$. An image is a set of configurations that *appear* the same to an ideal observer. Images inherit bonds from the configurations contained in them and can be combined (if bonds fit) and be transformed by similarities. $\mathcal{I}$ turns out to be a partial universal algebra with combinatory operations, with congruences and homomorphisms familiar to the algebraist.

The quotient space $\mathcal{P} = \mathcal{I}/\mathcal{S}$ is called the pattern family, its elements the *patterns* which are thought of as images modulo the invariances represented by the similarity group $\mathcal{S}$. The images are what can be observed by an ideal (with no loss of information) observer. The actual observer, however, may only be able to see the elements of the image algebra with loss of information due to observational noise or limited accuracy in the sensor. Denote the operation by which a pure image I appears as some object, say $I^{\mathcal{O}}$, by deformations

$$d\colon \mathscr{P} \to \mathscr{P}^{\mathscr{P}}, \qquad d \in \mathscr{D}. \tag{4}$$

Here $\mathscr{D}$, the *deformation mechanism*, can be random or deterministic.

2.1. *Regular Structures*

The *regular structures* $\mathscr{L}(\mathscr{R})$, $\mathscr{I}$, $\mathscr{P}$ are rigid constructs that represent knowledge about the 'typical' appearance of the phenomena under study. Often, however, variability is as important as typical appearance. Variabilities shall be formalized by introducing an *acceptor function* $A(\ ,\ )\colon \mathscr{B} \times \mathscr{B} \to R^+$ with an associate probability density p over $\mathscr{L}(\mathscr{R})$

$$p(c) = \frac{1}{Z} \prod_{\sigma} A\left[\beta_{j'}\{g(i')\},\ \beta_{j''}\{g(i'')\}\right]. \tag{5}$$

In equation (5) the product is taken over all segments $(i', j') - (i'', j'')$ in the graph σ with the constant Z normalizing p to have integral 1. A reader familiar with statistical mechanics recognizes A as $\exp(-E)$, E energy, and Z as the partition function.

The resulting probabilistic regularity is written as $\mathscr{R}' = \langle \mathscr{G}, \Sigma, A \rangle$. If the implication

$$\{A(\beta', \beta'') > 0\} \Rightarrow \rho(\beta', \beta'') \tag{6}$$

holds, then the support of the probability measure P induced by $\mathscr{R}'$ is contained in $\mathscr{L}(\mathscr{R})$. This means that almost surely all configurations will be regular. In the opposite case we shall, with some positive probability, encounter irregular configurations and we speak of *relaxed* (as opposed to rigid) regularity.

In the first case the probability measure P induces another probability measure on $\mathscr{I}$ through the natural map $\mathscr{L}(\mathscr{R}) \to \mathscr{I}$ belonging to the identification rule R. This allows us to ask the same questions for *probabilities on the image algebra* that classical probability theory has studied and to a great extent answered on the familiar algebraic structures **Z**, **R** and its Cartesian powers as well as other groups, vector spaces and topological algebras. For example, can we prove analogues of the law of large numbers or of the central limit theorem but on the regular structures that have typically only partial combinatory operations? Or, what are the properties of Markov processes taking values in $\mathscr{L}(\mathscr{R})$ or $\mathscr{I}$?

Such problems are studied in *metric pattern theory*, an emerging field in an incomplete stage of development. Many of the questions of inference in regular structures lead naturally to issues in metric pattern theory. A reader can find a presentation of it (at 1980) in Grenander (1981), chapter 5.

2.1.1. *Fixed and multiple-graph deformable templates*

Within these regular structures we choose a particular configuration, call it the *template*,

$$c^0 = \sigma(g^0(1), g^0(2), \ldots, g^0(n)) \in \mathscr{L}(\mathscr{R}). \tag{7}$$

Applying similarities $s(i) \in \mathscr{S}$ to the generators $g^0(i)$ yields a new configuration, the *deformed template*,

$$c = \sigma(g(1), g(2), \ldots, g(n)) = \sigma(s(0)\, g^0(1), s(2)\, g^0(2), \ldots, s(n)\, g^0(n)) \tag{8}$$

which will not always be regular. Make the assumption, one which is more convenient than necessary, that the equation $s\,g(1)=g^0$ for given $g(1)$, $g^0\in\mathcal{G}$ has a unique solution in s implying that for a fixed template c^0 the generators $g(i)$ are bijectively related to the similarities. It is then natural to introduce a prior measure P on the configuration space via a density on $\mathcal{S}^n$

$$p(s(1), s(2), \ldots, s(n))=\frac{1}{Z}\prod_{\sigma} A\{s(i'), s(i'')\} \tag{9}$$

with the product over all segments (i', i'') in the connector σ. If the deformed template configurations are required to be regular the probability must be conditioned on the set $\mathcal{L}(\mathcal{R})$. The result is a probability measure defined on $\mathcal{L}(\mathcal{R})$ which becomes the basis for the priors used throughout.

In this knowledge representation the template, or sometimes templates, express typical structure, with A describing the variability around it. In most applications of this model the similarity group $\mathcal{S}$ has been a low dimensional Lie group, but the product group $\mathcal{S}^n$ is of high dimension.

Thus far only templates associated with fixed graphs σ have been defined. Now extend the notion of deformable templates to include configurations on the family of graphs Σ, with an associated family of templates $\{c^0(\sigma)\}_{\sigma\in\Sigma}$. Inference in this greater knowledge representation involves deducing not only the $n(\sigma)$ similarity group values associated with each graph but the graph type $\sigma\in\Sigma$ as well. The full configuration space $\mathcal{L}$ becomes the union of configuration spaces over all graphs:

$$\mathcal{L}=\bigcup_{\sigma\in\Sigma}\mathcal{L}(\sigma). \tag{10}$$

Such a setting is essential for the family of *random graph* problems such as in *computational linguistics–language understanding* (see Mark *et al.* (1992) for example) in which the graphs carry the semantic and syntactic information associated with the language string. The graphs are the *randomly branched trees* (Harris, 1963) associated with *context-free languages* (Chomsky, 1956, 1959; Grenander, 1967; Miller and O'Sullivan, 1992). The graph type and its associated structure are fundamental to the deduction.

This view is also fundamental to the object recognition setting in which the number of objects are not known *a priori*. Graphs in Σ_{scene} are multiple-object scenes: the original templates c^0 become the nodes of the scene graphs σ_{scene}. We illustrate via the following example.

2.1.2. *Example: membranes and mitochondria*

The geometric shapes are generated via deformations of the one-dimensional template manifolds corresponding to *lines* and *circles*. The generators are induced by the tangent vectors of the template, directed arcs in $\mathbf{R}^2$ linearly connected. Each generator carries two bonds with values the end points of the directed arc segments. The bond relation $\rho=$ TRUE implies connectivity. The membranes are unclosed, and correspond to the graph type σ linear, with the mitochondrial sections closed implying that the first and last generator are connected in the graph σ cyclic.

The cell membranes are unclosed curves in the plane transformations of linear

templates, with mitochondria closed curves which are transformations of circular templates:

$$f_{\text{linear}}(m)=\int_0^m \begin{pmatrix} u(l) & v(l) \\ -v(l) & u(l) \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \mathrm{d}l+\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}, \tag{11}$$

$$f_{\text{cyclic}}(m)=2\pi\int_0^m \begin{pmatrix} u(l) & v(l) \\ -v(l) & u(l) \end{pmatrix}\begin{pmatrix} -\sin(2\pi l) \\ \cos(2\pi l) \end{pmatrix} \mathrm{d}l+\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}. \tag{12}$$

Closure on the mitochondria corresponds to $f_{\text{cyclic}}(0)=f_{\text{cyclic}}(1)$.

The continuum of similarities

$$s(l)=\begin{pmatrix} u(l) & v(l) \\ -v(l) & u(l) \end{pmatrix}\in\mathscr{S}(2)\equiv\text{uniform scale}\times\text{rotation}, \qquad (u, v)\in R^2\setminus\{0\},$$

are mapped to n(linear) or n(cyclic) unique operators $\{s(k)\}_{k=0}^{n-1}$ with $s(l)=s(k)$, $l\in(k/n, (k+1)/n)$, taking their action on the generators which are piecewise straight tangent segments to the line and circle. Note that the closure constraint forces the highest frequency discrete Fourier transform of the (u, v)-process to be 0 (see equation (24) later).

The parameter vector $x(\sigma)$ associated with configuration $c(\sigma)$ is the set of $n(\sigma)$ parameters encoding the global translation and locally applied scale and rotation. The closure constraint on the cyclic configurations reduces the dimension by 2 implying $x(\text{cyclic})\in\mathscr{G}(\text{cyclic})\equiv R^{2n(\text{cyclic})-2+2}$ whereas $x(\text{linear})\in\mathscr{G}(\text{linear})\equiv R^{2n(\text{linear})+2}$. As the scenes are unions of multiple objects, the number (call it m) of objects not known beforehand, the second graph family Σ_{scene} consists of multiple objects, $m=0, 1, \ldots$. The scene graphs $\sigma=\text{mult}(m_1, \text{linear})$ and $\sigma=\text{mult}(m_2, \text{cyclic})$ are the disconnected union of m_1 and m_2 linear and cyclic graphs, with the generators of scene graphs the objects themselves. The set of scene graphs becomes

$$\Sigma_{\text{scene}}=\bigcup_{m_1\geq 0}\text{mult}(m_1, \text{linear})\times\bigcup_{m_2\geq 0}\text{mult}(m_2, \text{cyclic}). \tag{13}$$

The parameter vector $x\{\text{mult}(m)\}\in\mathscr{G}\{\text{mult}(m)\}$ associated with $c\{\text{mult}(m)\}$, an m-object scene, is just the concatenation of parameters associated with each of the m objects and is of dimension

$$n\{\text{mult}(m)\}=\sum_{i=1}^{m} n\{\sigma(i)\}, \qquad \sigma(i)\in\{\text{linear, cyclic}\}.$$

The full configuration space becomes the union of configuration spaces $\mathscr{G}=\bigcup_{\sigma\in\Sigma_{\text{scene}}}\mathscr{G}(\sigma)$.

2.2. *Search: Diffusion and Jumps*

To carry out inference in these spaces it has become apparent that a powerful method, both for analytical study and computational implementation, is SDEs of the form

$$\mathrm{d}s(t)=\text{grad}\,\log p(s)\,\mathrm{d}t+\sqrt{2}\,\mathrm{d}W(t), \tag{14}$$

where p is the density on configuration space. Here $s=(s(1), s(2), \ldots, s(n))\in\mathscr{S}^n$ and W is the Wiener process with independent and identically distributed components. The solution of the equation is restricted to the set

$$\mathscr{M}=\{s|\sigma(s(1)g^0(1), s(2)g^0(2), \ldots, s(n)g^0(n))\in\mathscr{C}(\mathscr{R})\}. \tag{15}$$

The equation simulates a diffusion process on some manifold $\mathscr{M}$ and $s(t)$ moves along some continuous trajectory. The resulting configuration process $c(t)\in\mathscr{M}$ can be thought intuitively as the result of infinitesimal random transformations carried out in the group $\mathscr{S}^n$.

The existence of this multiple-graph space $\mathscr{C}$ motivates the introduction of a second transformation type on the templates, extending from such continuous transformations to discontinuous transformations. These transformations act by changing the graph type associated with the configuration to a new graph type and a new resulting configuration. These we term *simple graph moves*. The simple moves are drawn probabilistically from a family $\mathscr{F}$ of changes in the connector σ and are applied discontinuously, with the simple moves defining transitions through Σ, $\mathscr{F}$: $\Sigma\to\Sigma$. The family of graph transitions is chosen sufficiently large to act transitively in the sense that given any pair σ', $\sigma''\in\Sigma$ it should be possible to find a finite chain of transitions that leads from σ' to σ''. The set $\mathscr{F}$ controls the jump dynamics in the jump–diffusion processes described later.

2.3. *Inference in Pattern Theory*

The tasks of inference for regular structures take many forms with four mentioned here.

(a) The least challenging task, but one that has received most attention, is *image restoration*. Having observed a deformed image $I^{\mathscr{D}}=\mathrm{d}I$, find a pure image I^* approximating the true image I as well as possible in some specified sense. To the statistician this appears as a problem in (point) estimation; to the communication engineer a problem of filtering. From both points of view it is a familiar task with a formidable arsenal of methods existing of potential applicability.

If a probability measure P over the image algebra $\mathscr{I}$ is available then it is interpreted as prior probabilities in the Bayesian setting. This will be the case in the main body of the paper, but for many pattern theoretic inferences we must do without any prior, perhaps without any randomness whatsoever.

(b) Another task that has also generated many publications is *pattern recognition*. Given an observed image $I^{\mathscr{D}}$, decide to what pattern class in $\mathscr{P}$ it belongs. This can be viewed as testing statistical hypotheses or as a multiple-decision problem. Since the formal notion of pattern is based on congruence modulo the similarity group $\mathscr{S}$, the pattern recognition task leads to invariant decision procedures.

It is obvious that these first two tasks are closely related. When restricted to pictorial (geometric) image ensembles they have often been studied starting from what the neurophysiologists and psychologists know about visual systems in biology, emphasizing the distinction between low level (local) and high level (global) vision; this opposition will also appear later.

(c) A more ambitious task is *image understanding*. Take as an example a cytologist studying a micrograph, trying to understand what is seen in terms of the awesome body of knowledge that is today available about various organelles and other cell structures. Identify components, relate them to each other, make statements about the fine structure as well as the overall appearance of the cell. Or imagine a pathologist examining a slide, looking for deviations from the normal, requiring knowledge not only about the normal and variations around it but also information concerning the deviations that may occur and how likely they are.

To paraphrase this in more theoretical terms image understanding tasks come in two forms:

(i) *internal understanding* in which an observed image is explained within the domain of normal regular structures and
(ii) *external understanding* where the aim is to discover abnormalities that are not consistent with the model of normal regular structures.

Human observers do this well, although the task is time consuming and sometimes boring. For example PAP smears in gynaecology are examined by technicians, calling on the trained pathologist when required. Many attempts have been made to automate the inspection, so far with limited success.

To achieve even partial automation of image understanding of such complex images is a formidable task, that in our view requires the formalization of the subject-matter knowledge on which the biomedical expert relies. This is vaguely reminiscent of knowledge engineering in artificial intelligence (AI), but what we are attempting to do differs from AI in an important aspect. AI tries to imitate *general* human intelligence—we want to achieve algorithmic understanding only in a very *special* universe. To achieve this limited goal we believe that knowledge representation by regular structures offers a methodology that is sufficiently powerful and practically attainable with current computer technology.

(d) A fourth task, essentially different from the above, is to *create the regular structures* from the knowledge that is available, from data and experiments. How shall generators and bond relations be discovered, and how should the acceptor functions from observed images be estimated? What connection types and similarity groups should be chosen? Some of these questions can be answered to a partial extent. For example, if the regular structure has been chosen except for parameters in the acceptor function then we are dealing with an empirical Bayes problem, with standard methods available for its solution. However, the creation of the generator space has in most cases been done intuitively and only in isolated instances has it been possible to develop constructive methods that can be made into algorithms.

2.4. *Related Work*

The ideas of pattern theory were proposed by Grenander in 1976 at a scientific meeting in Loutraki, Greece. This was only a research programme and the analytical developments during the next decade had few applications. From 1980 onwards applications began to appear, at first only to image processing of local type.

This involved priors of the Ising type, but physicists had long employed such models to study ferromagnetism and other areas exhibiting critical behaviour. For this they had developed simulation techniques, as in Fosdick (1963), using variations of the Metropolis algorithm (Metropolis *et al.*, 1953).

In 1974 appeared the pioneering paper of Besag opening the way for further advances. Besag (1974) approached the problems from a different perspective—that of spatial statistics—but has much in common with the topics discussed here. They are also correlated with the ideas in the fundamental monograph Bartlett (1975) as well as the thought-provoking paper of Whittle (1954). The earliest attempt to introduce probabilistic couplings via graphs that we have been able to locate is Wright (1921) in his path diagrams.

Early attempts to use the MRFs on simple regular structure for describing textures, e.g. Horn (1977), were not encouraging. The computing resources were still quite limited and, more importantly, the generator spaces used were too naïve. Fresh insight occurred in Geman and Geman (1984), where more structured generators were introduced for describing edges. This study stimulated many others carried out in a similar spirit. In Grenander (1983) other regular structures were proposed and investigated but only using simulated data.

At about the same time it was suggested that global properties of images be explored by pattern theoretic knowledge representations, namely the probabilistically deformed templates. They had been suggested already in Freiberger and Grenander (1969) and Grenander (1970, 1985) but at that time available computer power was not sufficient for practical implementation. It took until the mid-1980s before this could be achieved: Knoerr (1988), Grenander *et al.* (1990) with the latter also aiming at automatic detection of abnormalities in the pattern, as well as Ripley (1988).

This is similar in spirit to the *physically based* modelling work of Terzopoulos and Waters (1990) on computerized surface models, the boundary finding work of Staib and Duncan (1992) and the elegant deformable CT work of Bajcsy and Kovacic (1989).

A somewhat different approach is Kendall's (1977, 1984) *shape theory*, which is also related to Bookstein (1978). Many ideas in morphometrics, and shape in general, can be traced back to the legendary work of Thompson (1917), especially the famous last chapter.

Developments in computer architecture have extended the practical scope of these statistical techniques. An obvious instance of this is the advent of accessible massively parallel SIMD machines, and more recently the promising computational methods based on analogue devices, transistor based as in Mead (1989) or employing charge-coupled devices as in Wyatt (1992).

The interaction between analytical advances and hardware innovations may lead to hybrid architectures, perhaps directly coupled to the sensors by fast networks. An important statistical spin-off from the work on MRFs was the computational methods for obtaining maximum *a posteriori* estimates. This was done by simulated annealing and analytical conditions were derived for convergence of the algorithm: Geman and Geman (1984), Gidas (1993) and Hajek (1988). Perhaps more relevant to the analogue and SIMD computational methods has been the work in SDE search. Early on Grenander (1983) proposed Langevin's equations for simulating distributions, and more recently Gidas (1993), Geman and Hwang (1987) and Amit *et al.* (1991). In Miller *et al.* (1991) and Roysam and Miller (1992) these ideas have been applied to Gibbs distributions on discrete spaces for hypothesis testing and symbolic inference,

with computational methods discussed for SIMD machine and analogue implementation. For substantive reviews of random sampling for Bayesian inferences see Besag and Green (1993).

Concerning the combination of jump and diffusion dynamics, to our knowledge Feller (1936) was the first to describe the Kolmogoroff backward and forward dynamics of Markov processes having jump and diffusion parts within a single subspace. We have drawn heavily on results from Ethier and Kurtz (1986), p. 266, in which Markov processes with diffusions are constructed over countable subspaces.

3. JUMP-DIFFUSION RANDOM SAMPLING OF THE POSTERIOR

Having defined the configuration space $\mathscr{G} = \cup_{\sigma \in \Sigma} \mathscr{G}(\sigma)$ as the union of spaces over which the inference is to be performed, the crucial part of the problem still remaining is the derivation of the inference algorithm for choosing the graphs and their associated transformations, i.e. *how to carry out hypothesis formation*. We are not asking for *ad hoc* answers to this question but shall try to deduce the algorithm from the regular structure expressing our prior knowledge on the configuration space.

The method of inference is to construct a single posterior distribution over $\mathscr{G}$ and then to sample from it via a Markov process $X(t)$ which satisfies *jump–diffusion dynamics*. For this, model k and the real Euclidean space $R^{n(k)}$ are identified with a particular graph $\sigma \in \Sigma_{\text{scene}}$ of dimension $n(k)$. The full hypothesis space becomes a countable union of Euclidean spaces $\mathscr{G} = \cup_{k=0}^{\infty} R^{n(k)}$. The posterior distribution μ is a Gibbs distribution over the collection of spaces, i.e., for all measurable $\mathscr{A}$

$$\mu(\mathscr{A}) = \frac{\sum_{k=0}^{\infty} \int_{\mathscr{A} \cap R^{n(k)}} \exp\{-H_k(x)\}\,dx}{Z} \tag{16}$$

with Gibbs density $\exp\{-H_k(x)\}/Z$, $x \in R^{n(k)}$, and dx Lebesgue measure appropriate for the space. The Markov process $X(t)$ with sample paths which are the inferences is said to satisfy *jump–diffusion* dynamics through $\mathscr{G}$ in the sense that

(a) on random exponential times the process jumps from one of the countably infinite set of spaces $R^{n(k)}$, $k = 0, 1, \ldots$, to another and
(b) between jumps it satisfies SDEs that are of dimension appropriate for that space.

The proper choice of jump and diffusion parameters make μ on $\mathscr{G}$ invariant. From this it follows that ergodic averages generated from the process converge to their expectations. These results are now stated as two theorems.

Theorem 1. If the jump–diffusion process $X(t)$ with state space $\mathscr{G} = \cup_{k=0}^{\infty} R^{n(k)}$ has the properties that

(a) the diffusion $X(t)$ within any subspace $R^{n(k)}$ satisfies the stochastic differential equation

$$dX(t) = -\frac{1}{2} \nabla H_k\{X(t)\}\,dt + dW_{n(k)}(t) \tag{17}$$

with $X(t)$, $\nabla(\,)$ and $W_{n(k)} \in R^{n(k)}$ the state, gradient and standard vector Brownian motion respectively, with the gradient $\nabla(\,)$ satisfying Lipschitz continuity, and

(b) the jump intensities and transition probability $q(x, dy)$, $q(x)$ and $Q(x, dy)$ defined in the standard way (Gihman and Skorohod, 1965) with $q(x) = \int_{\mathscr{G} \setminus x} q(x, dy)$ and $Q(x, dy) = q(x, dy)/q(x)$ bounded continuous functions with the jumps local satisfying

$$q(x)\,\mu(dx) = \int_{\mathscr{G}} q(y)\,Q(y, dx)\,\mu(dy), \tag{18}$$

then $X(t)$ is a Markov process on $\mathscr{G}$ with invariant measure μ.

(The jumps are *local* in the sense defined in Amit *et al.* (1993), i.e. there is some constant B such that the distance between where the jump process jumps from to where it jumps to is less than B. This ensures that the domain of the semigroup is $\hat{C}(\mathscr{G})$ the closure of $C_c^2(\mathscr{G})$.)

Theorem 2. Let $X(t)$ be the Markov process satisfying theorem 1. Assume that the Euclidean spaces are connected under the jumps, i.e. $\forall k, k'$, $\exists j(k, k')$ finite sequence of simple graph moves carrying the process from $R^{n(k)}$ to $R^{n(k')}$.

Then μ is the unique invariant measure of the jump–diffusion process $X(t)$, and for all $x \in \mathscr{G}$ the associated chain $X(i\Delta)$, $\Delta > 0$, converges in total variation norm to μ the invariant measure.

Proof. The proof of theorem 1 relies on the fact that the generator, or backward Kolmogoroff operator A for the jump–diffusion process, characterizes the stationary measure, i.e. μ is stationary for $X(t)$ if and only if $\int A f(x)\,\mu(dx) = 0$ for all f in a large family, the *core* of A (Ethier and Kurtz (1986), p. 239, and the Echeverria theorem, Ethier and Kurtz (1986), p. 248). Now the generator is the superposition of the diffusion and jump generators $A = A^d + A^j$ (diffusion plus jump), both standard. This follows from Ethier and Kurtz (1986), p. 266. The core $C_c^2(\mathscr{G})$ is the set of functions

$$f(x) = \sum_{k=0}^{m} \mathbf{1}_{R^{n(k)}}(x)\, f_k(x), \qquad m \geqslant 0,$$

and $f_k(x) \in C_c^2(R^{n(k)})$ twice continuously differentiable compactly supported functions on $R^{n(k)}$. Applying A to such f and integrating with respect to μ gives

$$\int_{\mathscr{G}} A f(x)\,\mu(dx) = \frac{1}{2}\sum_{k=0}^{m} \int_{R^{n(k)}} \left[-\sum_{i=1}^{n(k)} \frac{\partial H_k\{x(k)\}}{\partial x_i} \frac{\partial f_k\{x(k)\}}{\partial x_i} + \sum_{i=1}^{n(k)} \frac{\partial^2 f_k\{x(k)\}}{\partial x_i^2} \right] \frac{\exp[-H_k\{x(k)\}]}{Z}\, dx(k)$$

$$+ \int_{\mathscr{G}} \mu(dx)\, q(x) \left[\int_{\mathscr{G}} \{f(y) - f(x)\} Q(x, dy) \right], \tag{19}$$

with the first part the standard SDE operator and the second the jump operator. To show that $\int A f(x)\,\mu(dx) = 0$, integrate by parts once the second derivative term in the first part of equation (19) and use the second condition (b) in the theorem statement.

Proving theorem 2 that the process $X(i\Delta)$ converges to μ is a result of irreducibility which has two parts. First the SDE within each subspace is irreducible over compact sets (by boundedness of the drift coefficients over compact sets (Grenander and Miller, 1991)). Connectedness of the different spaces via the graph moves gives irreducibility over the entire space (Grenander and Miller, 1991). To see that the Markov chain

has a unique invariant measure follows from the fact that the process is irreducible with bounded invariant measure μ (theorem 1), implying that the chain is Harris recurrent with measure μ and that μ is the only invariant probability measure (Revuz, 1975). Recurrence and the existence of the invariant measure μ establishes the variation norm convergence for the associated chains (Athreya and Ney, 1978). □

The particular choice of jump dynamics which will satisfy condition (b) of theorem 1 still has to be specified. Reversibility of the graph moves will be required. For this define $\mathscr{T}^1(k) \subset \Sigma$ to be the set of subspaces which are reachable in one graph move from k, and $\mathscr{T}^{-1}(k) \subset \Sigma$ the set from which k can be reached in one move. Also define $\mathscr{L}\{\mathscr{T}^1(k)\}$ and $\mathscr{L}\{\mathscr{T}^{-1}(k)\}$ to be the unions of subspaces in $\mathscr{L}$ associated with $\mathscr{T}^1(k)$ and $\mathscr{T}^{-1}(k)$ respectively. We have used two strategies which make the distribution μ invariant.

The first has acceptance–rejection dynamics analogous to Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990) and is constructed from a set of times $w_1, w_2, \ldots$ independent and exponentially distributed, with jump times $t_0 = 0$, t_1 and t_2 defined according to

$$t_i = \inf\left\{t: \int_{t_{i-1}}^{t} q(X_s)\, ds \geqslant w_i\right\}.$$

The process $X(t)$ satisfies the SDE between jump times, with the process moving from one subspace to another on the t_is with transition probability measure $Q(x, dy) = q(x, dy)/q(x)$.

The second simulation method has jump dynamics of the acceptance–rejection Metropolis type (Metropolis *et al.*, 1953; Hastings, 1970). On each candidate jump time $t_i = \Sigma_{j=1}^{i} w_j$ a new candidate state is drawn from the prior. The state is deterministically accepted if the energy in the likelihood term decreases and is probabilistically accepted with probability exponential to the negative increase in energy.

The choices for the jump intensities corresponding to these procedures are given as follows.

Corollary 1. Assume that the jump–diffusion process satisfies part (a) of theorem 1, and reversibility on the graph moves $\mathscr{T}^1(k) = \mathscr{T}^{-1}(k)$.

If

$$q(x(k), dy(k')) = \mu\{dy(k')\} \qquad \text{for } y(k') \in \mathscr{L}\{\mathscr{T}^1(k)\}, \tag{20}$$

and is 0 otherwise, then μ is a stationary measure of the process.

Corollary 2. Assume that the jump–diffusion process satisfies part (a) of theorem 1, a reversibility condition on the graph moves $\mathscr{T}^1(k) = \mathscr{T}^{-1}(k)$ and the energy of the posterior distribution can be written as $H(x) = L(x) + P(x)$ (P is the potential associated with the prior). Define $[f]_+$ to denote the positive part of the function f.

If

$$q(x(k), dy(k')) = \exp(-[L\{y(k')\} - L\{x(k)\}]_+) \exp[-P\{y(k')\}]\, dy(k')$$
$$\text{for } y(k') \in \mathscr{L}\{\mathscr{T}^1(k)\} \tag{21}$$

and is 0 otherwise, then μ is a stationary measure of the process.

Proof. The required continuity and boundedness properties follow from the properties of the posterior. Part (b) of theorem 1 is proved for either corollaries by simply integrating the jump intensities. □

4. APPLICATION TO SUBCELLULAR ORGANELLES

For specificity of description fix the number of arcs in each graph at $n(\text{cyclic}) = n$ and assume that all graphs are cyclic. Then the mth model is identified with m-object scenes of dimension $n\{\text{mult}(m, \text{cyclic})\} = 2nm$. The parameter vectors are $x(m) \in R^{2nm}$ associated with an m-object scene mult$(c(1), \ldots, c(m))$, and the full configuration space $\mathscr{G} \equiv \cup_{m=0}^{\infty} R^{2nm}$. The posterior $\mu(dx)$ having Gibbs density $\exp[-H_k\{x(k)\}]/Z$, $x(k) \in R^{n(k)}$, is constructed as follows.

4.1. *Bayes Posterior for Electron Micrograph Shapes*

4.1.1. *Prior distribution*

The prior distributions on the single objects lay at the heart of Bayesian inference. To incorporate notions of curvature, the (ρ, θ)-parameters parameterizing the scale–rotation groups are assumed *stationary*. For the cyclic curves they are also *circulant*, implying the 2×1 vectors $z(k) = (u(k) = \rho(k)\cos\theta(k),\ v(k) = \rho(k)\sin\theta(k))^{\mathrm{T}}$ have a block circulant, Toeplitz covariance

$$\mathbf{K} = \begin{pmatrix} K(0) & K(1) & K(2) & \ldots & K(n-1) \\ K(n-1) & K(0) & K(1) & \ldots & K(n-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K(1) & K(2) & \ldots & \ldots & K(0) \end{pmatrix}. \tag{22}$$

The $K(j) = E\{z(k)z(k+j)^{\mathrm{T}}\}$ are 2×2 blocks. The prior is chosen to be Gaussian on the zs. We block diagonalize $\mathbf{K}$ with the block DFT matrix. The rotated variables $\{\bar{z}(k) = (\bar{u}(k), \bar{v}(k))^{\mathrm{T}}\}_{k=1}^{n(\text{cyclic})}$ where $\bar{u}(k) = \Sigma_{i=0}^{n-1} u(i)\, W(i, k)$, $\bar{v}(k) = \Sigma_{i=0}^{n-1} v(i)\, W(i, k)$ are used, with

$$W(i, k) = \frac{1}{\sqrt{n}} \exp\left(-\mathrm{j}\frac{2\pi ik}{n}\right) \qquad (\mathrm{j} = \sqrt{(-1)}). \tag{23}$$

Then, with $\bar{z}(0)$ and $\bar{z}(n/2)$ real Gaussian 2×1 vectors and $\{\bar{z}(k)\}_{k=1}^{n/2-1}$ complex Gaussian 2×1 vectors of Goodman (1963) class all independent with covariance

$$\Lambda(k) = \sum_{i=0}^{n-1} K(i)\, W(i, k)$$

and Hermitian symmetry $\bar{z}(n-k) = \bar{z}^*(k)$, the $z(k)$ are real Gaussian with block covariance $\mathbf{K}$. If the (u, v)s are in turn uncorrelated, so that the covariance matrices $K(k)$ are diagonal, then $\Lambda(k)$ is diagonal. This is not the case.

The closure constraint on the simple curves $f(0) = f(1)$ of equation (12) is straightforwardly incorporated into this representation since

$$\sum_{k=0}^{n-1}\{u(k)-\mathrm{j}v(k)\}\left[\exp\left\{\mathrm{j}\frac{2\pi(k-1)}{n}\right\}-\exp\left(\mathrm{j}\frac{2\pi k}{n}\right)\right]=0$$

implies that the highest frequency discrete Fourier transform coefficient of $u-\mathrm{j}v$ is 0:

$$\sum_{k=0}^{n-1}\{u(k)-\mathrm{j}v(k)\}\exp\left\{-\mathrm{j}\frac{2\pi k(n-1)}{n}\right\}=0, \tag{24}$$

giving $\bar{u}(n-1)=\mathrm{j}\,\bar{v}(n-1)$. Hermitian symmetry implies $\bar{u}^*(1)=\mathrm{j}\,v^*(1)$ giving

$$\Lambda(1)=\begin{pmatrix}\sigma^2 & \mathrm{j}\sigma^2\\ -\mathrm{j}\sigma^2 & \sigma^2\end{pmatrix}$$

where $\sigma^2=E\{\bar{u}(1)\,\bar{u}^*(1)\}$. The effect of closure is to reduce the dimension of the $2n$ (u, v)-variables to $2n-2$.

Remark 1. Strictly the boundaries should satisfy a global constraint: they should be non-self-intersecting. In HANDS (Grenander *et al.*, 1990) the solution to this was provided and shown that in practice this constraint can be neglected since the probability of self-intersection is negligible for parameter values of interest.

4.1.2. *Likelihood and pseudolikelihood*

The underlying ideal image $I=(c(1), c(2), \ldots)$, or truth, consists of the lists of parameters encoding the global translations and local scales and rotations. The measurement process senses the data $I^{\mathscr{D}}$. The conditional density $p(I|I^{\mathscr{D}})\propto p(I)\,L(I^{\mathscr{D}}|I)$ relates the ideal I to the data $I^{\mathscr{D}}$, with $L(I^{\mathscr{D}}|I)$ and $p(I)$ the sensor likelihood model and the prior on the ideal respectively.

The driving force for the shape models comes from the connection between shapes and the pixel data as follows. The data $I^{\mathscr{D}}$ are assumed to be a superposition of random shapes with random interiors $\mathscr{Y}$ in the finite domain $\mathscr{S}\subset R^2$. Each interior is a realization from a Gibbs random field density associated with mitochondria (mito), membrane (mem) and cytoplasm (cyt):

$$p_\theta(\mathscr{Y})=\frac{\exp\{-E_\theta(\mathscr{Y})\}}{Z_\theta}, \qquad \mathscr{Y}\subset\mathscr{S}, \tag{25}$$

$\theta\in\{\text{mito, mem, cyt}\}$. Choosing object $c(j)$ to be model $\theta(j)$ implies that its interior $\mathscr{Y}_{c(j)}$ has Gibbs potential $E_{\theta(j)}$, $\mathscr{S}\setminus\cup_j\mathscr{Y}_{c(j)}$ having potential E_{cyt}. Assuming that objects are independent then

$$L\{I^{\mathscr{D}}|I=(c(1), c(2), \ldots)\}=\left[\prod_j\frac{\exp\{-E_{\theta(j)}(\mathscr{Y}_{c(j)})\}}{Z_j}\right]\frac{\exp\{-E_{\text{cyt}}(\mathscr{S}\setminus\cup_j\mathscr{Y}_{c(j)}\}}{Z}. \tag{26}$$

In general the usual difficulty is met in computing the partition functions associated with the random sets. We take several approaches. The first two involve the essential specification of the random field via its conditional probabilities, thereby avoiding the partition function. For this we have

(a) assumed independent pixel model probabilities and alternatively
(b) used Besag's *pseudolikelihood* with the conditional probabilities and the neighbourhood sizes estimated from the micrograph data using Smith and Miller's (1989, 1990) results on minimum description length estimation for random fields.

In both cases, the log-probabilities are continuously interpolated to a potential $E_\theta(y)$ to cover the image field $\mathcal{S}$ with a zero boundary added to all micrographs. Each of the Gibbs terms in the product of equation (26) is then of the form $\exp\{-\int_{\mathcal{Y}_{c(j)}} E_{\theta(j)}(y)\,dy\}$. The two approaches give similar results, and we show only those computed with pseudolikelihood in this paper.

The resulting posterior density on the parameters $x(m)$ associated with a scene mult($c(1), c(2), \ldots, c(m)$) in the m-object space is of the Gibbs form with potential

$$H_m\{x(m)\} = \sum_{j=1}^{m}\left[\int_{\mathcal{Y}_{c(j)}} E_{\theta(j)}(y)\,dy + E_{\sigma(j)}\{c(j)\}\right] + \int_{\mathcal{S}\setminus\cup_j \mathcal{Y}_{c(j)}} E_{\text{cyt}}(y)\,dy. \tag{27}$$

Here E_σ, $\sigma\in\{\text{linear, cyclic}\}$ denotes the potential of the prior on single objects. To ensure that the posterior integrates over the multiple object spaces, a Poisson prior on object number with mean estimated from the micrographs is added.

Remark 2. A third fully Bayesian method is being investigated which incorporates neighbourhood dependence. Assume that there is a single texture and that the intensities are $x(i)=\phi\{y(i)\}$ where ϕ is a fixed monotonically increasing function and $y(i)$ forms a stochastic process with means $m(i)$, $y(i)=z(i)+m(i)$. The random y-field should satisfy a partial stochastic difference equation $(Lz)(i)=e(i)$ where $e(i)$ is a Gaussian white noise process $N(0, 1)$. The difference operator L is to be non-singular with boundary chosen so that the problem is well posed in the usual sense. The joint density of the z-process is Gaussian with quadratic form $\|Lz\|^2$.

Since we have two textures we assume $(m_{\theta(1)}, L_{\theta(1)})$ and $(m_{\theta(2)}, L_{\theta(2)})$ over the two subsets of pixels which the hypothesis divides the picture. The likelihood is proportional to

$$\exp\left\{-\frac{1}{2}\left(\sum_1 [L_{\theta(1)}\{y(i)-m_{\theta(1)}(i)\}]^2 + \sum_2 [L_{\theta(2)}\{y(i)-m_{\theta(2)}(i)\}]^2\right)\right\},$$

with the sums extended over respective subsets of the picture.

Remark 3. In practice the posterior is modified to impose the constraint that in electron micrographs two objects cannot be superimposed, requiring an intersection penalty $\alpha\Sigma_{i,j}\int_{\mathcal{Y}_{c(i)}\cap\mathcal{Y}_{c(j)}} dy$.

4.2. *Jumps and Diffusions*

4.2.1. *Jumps and graph moves*

The family of connector graph changes $\mathcal{F}$ on $\Sigma\equiv\cup_{m=0}^{\infty}$mult($m$, cyclic) will determine which of the jump measures are non-zero, as well as the connectedness

and reversibility. We define the graph changes $T \in \mathscr{T}$ to consist of the addition of one object at a time or deletion of one of the existing objects:

$$T_b\colon \text{mult}(m) \xrightarrow{\text{birth}} \text{mult}(m+1),$$
$$T_d\colon \text{mult}(m) \xrightarrow{\text{death}} \text{mult}(m-1).$$

Jumps in parameter space take the form

$$x(m) \in R^{2nm} \xrightarrow{\text{birth}} y(m+1) \in R^{2n(m+1)},$$
$$x(m) \in R^{2nm} \xrightarrow{\text{death}} y(m-1) \in R^{2n(m-1)}.$$

A special notation is used to denote the location in the list of the added or deleted parameters. Define $x(m) \oplus_j y(1)$ to be the configuration generated from the birth of a new object $y(1) \in R^{2n}$ into the jth location in the list and $x^{(j)}(m) \in R^{2n(m-1)}$ the jth object removed from the list. $j = 1, \ldots, m+1$ birth changes in the graph mult(m) are allowed and m possible death changes implying that the transition measures from space R^{2nm} have mass on spaces $R^{2n(m+1)}$ and $R^{2n(m-1)}$, and are singular with respect to Lebesgue measure in each subspace:

$$\begin{aligned} q\{x(m),\, dy(m+1)\} &= \sum_{j=1}^{m+1} q_b\{x(m),\, y(m+1)\}\delta_{x(m)}\{dy^{(j)}(m+1)\}\, dy(1) \\ q\{x(m),\, dy(m-1)\} &= \sum_{j=1}^{m} q_d\{x(m),\, y(m-1)\}\delta_{x^{(j)}(m)}\{dy(m-1)\}. \end{aligned} \tag{28}$$

This choice $\mathscr{T}$ of graph changes implies that the jump intensity is

$$q\{x(m)\} = \sum_{j=1}^{m+1} \int_{R^{2n}} q_b\{x(m),\, x(m) \oplus_j y(1)\}\, dy(1) + \sum_{j=1}^{m} q_d\{x(m),\, x^{(j)}(m)\}. \tag{29}$$

4.2.2. *Diffusions*

There is an SDE for each of the spaces R^0, R^n, R^{2n}, . . .

$$dX(t) = -\frac{1}{2} \nabla H_m\{X(t)\}\, dt + dW_{2nm}(t), \tag{30}$$

$m = 0, 1, \ldots$, with $X(t)$, $\nabla H_m\{X(t)\}$ and $W_{2nm} \in R^{2nm}$, consisting of m sets of $2n \times 1$ vectors. The drifts are variations of the Gibbs posterior energy with respect to the scale, rotation and translation parameters of each of the m objects. These are obtained by viewing the curve family of equation (12) as single one-parameter curves $f(\gamma)$, $\gamma \in \{u(k), v(k), (x_0, y_0)\}_{k=1}^{n}$, with interior $\mathscr{D}_{f(\gamma)} \subset R^2$. The drifts are variations of the Gibbs potential

$$E(\gamma) = \int_{\mathscr{D}_{f(\gamma)}} E_{\text{int}}(y)\, dy + \int_{\mathscr{S} \setminus \mathscr{D}_{f(\gamma)}} E_{\text{ext}}(y)\, dy,$$

(E_{int} associated with interior) which reduces to a curvilinear integral along the boundary.

Theorem 3. For curves given by equation (12) with continuous potentials E_{int} and E_{ext}, then

$$\frac{\partial E(\gamma)}{\partial \gamma} = \int_0^1 J(l;\ \gamma)\,[\,E_{\text{int}}\{f(l)\} - E_{\text{ext}}\{f(l)\}\,]\ \mathrm{d}l, \tag{31}$$

$\gamma \in \{u(k),\ v(k),\ (x_0,\ y_0)\}_{k=1}^{n(\text{cyclic})}$, with the Jacobian determinants

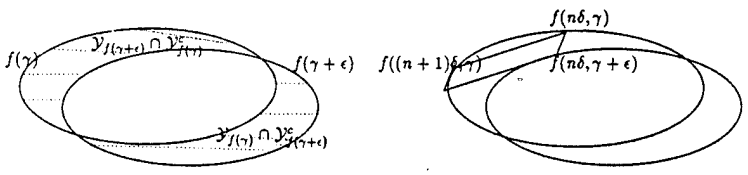
$$\begin{aligned} J\{l;\ u(m)\} &= -2\pi\left\{\sin(2\pi m)\frac{\partial f_y(l)}{\partial l} + \cos(2\pi m)\frac{\partial f_x(l)}{\partial l}\right\}\mathbf{1}_{\geqslant m}(l), \\ J\{l;\ v(m)\} &= 2\pi\left\{-\cos(2\pi m)\frac{\partial f_y(l)}{\partial l} + \sin(2\pi m)\frac{\partial f_x(l)}{\partial l}\right\}\mathbf{1}_{\geqslant m}(l), \end{aligned} \tag{32}$$

and $J(l;\ x_0) = 2\pi\,\partial f_y(l)/\partial l$, $J(l;\ y_0) = -2\pi\,\partial f_x(l)/\partial l$, $\partial f_x(l)/\partial l = -\sin(2\pi l)\,u(l) - \cos(2\pi l)\,v(l)$, $\partial f_y(l)/\partial l = \cos(2\pi l)\,u(l) - \sin(2\pi l)\,v(l)$.

Proof. The proof is geometric in nature. For small variations of the energy $E(\gamma+\epsilon) - E(\gamma)$ the Gibbs potentials must be computed over the areas shown in Fig. 2(a). The integral over the area is asymptotically ϵ, $\delta \downarrow 0$ obtained as the limit of the sum of parallelograms whose sides are the vectors $f(n\delta,\ \gamma) - f(n\delta,\ \gamma+\epsilon)$ and $f\{(n+1)\delta,\ \gamma\} - f(n\delta,\ \gamma)$. The area is (with sign) given by the sign theorem in trigonometry by the Jacobian determinant. In the limit the gradient becomes the contour integral (see Grenander and Miller (1991)). □

The curvilinear integrals are computed assuming that the number of scale–rotation similarities are sufficiently dense that the difference in potential along the generators (piecewise straight arcs) is constant, reducing the integral to a sum. See Grenander and Miller (1991), theorem 5.1 and its corollary.

To {*close the curves*} the SDE is performed in the rotated space of the $n \times n$ linear transformation $\mathbf{W}$ of equation (23) according to $\bar{\mathbf{u}} = \mathbf{W}\mathbf{u}$ and $\bar{\mathbf{v}} = \mathbf{W}\mathbf{v}$. Closure is enforced by setting the coefficients $\bar{u}(n-1) = \mathrm{j}\,\bar{v}(n-1)$ and $\bar{u}^*(1) = \mathrm{j}\,v^*(1)$ according to equation (24), and diffusing through the $2n-2$ dimensional space of the other $(\bar{u},\ \bar{v})$-coefficients. Their drifts are obtained by rotating the gradients with respect to the scale–rotations from theorem 3:



(a) (b)

Fig. 2. (a) Curves $f(\gamma)$ and $f(\gamma+\epsilon)$ with the shaded areas depicting the areas over which the integrals must be computed for the difference in potentials for the two curves; (b) parallelogram used to approximate the area with sides $f(n\delta,\ \gamma) - f(n\delta,\ \gamma+\epsilon)$ and $f\{(n+1)\delta,\ \gamma\} - f(n\delta,\ \gamma)$

$$\nabla_{\bar{u}} H(\mathbf{u}, \mathbf{v}) = (1+\mathrm{j})\mathbf{W} + \nabla_u H(\mathbf{u}, \mathbf{v}),$$
$$\nabla_{\bar{v}} H(\mathbf{u}, \mathbf{v}) = (1+\mathrm{j})\mathbf{W} + \nabla_v H(\mathbf{u}, \mathbf{v}). \tag{33}$$

The $2n-2$ components in each of the m-objects of $\nabla H\{X(t)\}$ in the SDE of equation (30) are given by the first $n/2+1$ real and imaginary parts of $\nabla_{\bar{u}} H$ and $\nabla_{\bar{v}} H$, with the last two entries $\partial H(x_0)/\partial x_0$ and $\partial H(y_0)/\partial y_0$.

Theorem 4. If the SDE is as defined in equations (30), (31) and (33) and jump measures equations (28) and (29) with the values for q_b and q_d chosen analogous to corollary 1

$$q_b\{x(m), x(m) \oplus_j y(1)\} = \exp[-H_{m+1}\{x(m) \oplus_j y(1)\}],$$
$$q_d\{x(m), x^{(j)}(m)\} = \exp[-H_{m-1}\{x^{(j)}(m)\}], \tag{34}$$

or analogous to corollary 2

$$q^b\{x(m), x(m) \oplus_j y(1)\} = \frac{1}{2(m+1)} \exp(-[L\{x(m) \oplus_j y(1)\} - L\{x(m)\}]_+) \exp[-P\{y(1)\}],$$

$$q^d\{x(m), x^{(j)}(m)\} = \frac{1}{2m} \exp(-[L\{x^{(j)}(m)\} - L\{x(m)\}]_+)$$

with L the posterior with the Gaussian shape prior removed, then the jump–diffusion satisfies theorems 1 and 2.

Proof. Clearly the Jacobian determinants (31) have bounded derivatives with respect to the scale, rotations and translation parameters. That coupled with the assumption that the texture potentials are bounded and differentiable implies that the drift terms are Lipschitz (see Apostol (1974)). The jump parameters $q(x)$ and $q(x, \ )$ are bounded and continuous. We need only to show the second condition (b) of theorem 1 to prove that $\mu\{dx(m)\}$ is stationary. Our choice for q_b and q_d is the analogue of corollary 1 to theorem 1, but the fact that the jump measures are not absolutely continuous with respect to the underlying Lebesgue measures forces a separate proof of condition (b) of theorem 1. To show condition (b) of theorem 1, substitute the defined birth and death intensities into the left- and right-hand sides of equation (18). The jump moves do satisfy the reversibility condition $\mathscr{T}^{-1}(m) = \mathscr{T}^1(m)$, implying that μ is a stationary measure. It is unique since the connectedness conditions of $\Sigma = \cup_{m=0}^{\infty} \mathrm{mult}(m)$ are satisfied, and from theorem 2 it is the unique density and the process converges to it. □

Remark 4. The algorithm has also been implemented for linear membranes by using the unclosed curves. For this, gradients with respect to the similarities are obtained by modelling the unclosed membranes as structures of constant known width 2δ pixels, piecewise linear and of variable length much greater than δ. The similarities are constrained to have uniform scale, with length the number of generators n(linear). Perturbing the boundary wiggles the linear structure of constant width and variable length n(linear), with position snaking through R^2 and parameterized via its midcurve $f(l)$, $l \in [\mathbf{0}, n(\text{linear})]$. To compute the curvilinear integral around the boundary of the membrane it is divided into two major components $f_{\delta^+}(l)$ and $f_{\delta^-}(l)$

determined by the midcurve $f(l)$ and its normal $n(l)$: $f_{\delta+}(l)=f(l)+\delta n(l)$, $f_{\delta-}(l)=f(l)-\delta n(l)$. Then a simple formula arises for computing the variation with respect to the midcurve parameters under the assumption that the texture potentials are constant along the linear generators. See corollary 2 to theorem 5.1 in Grenander and Miller (1991).

The family of graph moves $\mathscr{F}$ has also been enlarged to include addition and deletion of arc generators, which is extremely important for the tracking examples shown. Changes are also allowed which open and close segments thereby changing cyclic to linear graphs and linear to cyclic graphs. Finally splitting and merging of cyclic objects are allowed.

5. RESULTS

5.1. *Computational Results on Single-instruction, Multiple-data Architectures*

The automated inference has been implemented on the SIMD distributed array processor (DAP) of active memory technology and the MASPAR of MasPar Corporation, the distinctive feature being that on a single cycle all processors perform the identical instruction operating on their own local data store. These machines are in a line of locally connected massively parallel SIMD processors whose earliest conception was given by Slotnick in 1962 (Gregory and McReynolds, 1963) offering processor numbers of between 4000 and 16000. With the advent of these inexpensive massively parallel processors it is possible to implement imaging algorithms with computation times that are several orders of magnitude lower than that obtained with conventional processors.

The implementation closely parallels the structure of the machine and divides into two basic components. The first is the generation of a virtual object array in which the *local gradient* computations associated with the curvilinear integrals for diffusing each object are generated, independently of other objects in the scene. The second are the *global* jump operations associated with aggregation of the pseudolikelihood texture statistics over the object interiors.

The key to the computations is the distributed representation of the conditional probabilities allowing for rapid fetching of the statistics. This is done by choosing the image array to have the same topological structure as the processors themselves, requiring the mapping of the 128×128 pixel image to 4 pixels per processor. Each processor is viewed as *smart memory* storing the series of nonparametric models determined by the set of MRF conditional probabilities. These conditional probabilities can be computed from local neighbouring processor values.

There are two very different computations required for the jump–diffusion algorithm, as described below.

5.1.1. *Local diffusion operation*

Envision a virtual object array containing a variable number of objects, each object a column of processors dedicated to the computation of the diffusion of the group parameters encoding that object in the scene. For the 64×64 processor the number of objects is limited to 64, with the longest boundary having 64 scale–rotation elements encoding it. The gradient computation for each object has both a communication and a computation component.

The {*communication*} is associated with each object fetching from the array the difference in likelihood statistics along the boundary. As part of this communication, the intersections between boundaries are computed for determination of the penalty associated with boundary overlap. This penalty prevents local crossing of adjacent boundaries. As the boundaries are linear and represented by the sample points at the ends of each arc, the necessary operations required are to grow *fences* between the sample vertex points.

The {*computation*} is associated with the series of scale–rotation parameters encoding each object with the Jacobian determinants which must be computed. This requires four multiply cycles per boundary element. All 64 arc elements and 64 boundaries are completed in four of these multiply cycles. The Jacobian determinants are computed in a 4×2 multiply cycle for all boundaries and group elements, four for the scales–rotations and positions, and two multiply cycles per determinant. The curvilinear integrals are completed for each boundary by doing the 64 multiply integrals for the 64 parameter sets per boundary in a single multiply cycle.

5.1.2. *Global computations for jump operations*

The family of graph changes consists of births of new objects in one of a finite number of places on the grid, deleting any of the existing objects, merging any of the objects and splitting the objects. The jump moves require aggregation of texture models over the entire array. *These are area integrals.* The masking and cellular automata nature of the distributed SIMD class of architectures is *ideal.*

To compute a jump move all boundary co-ordinates associated with an object are placed onto the array. The boundary is fenced, giving the contiguous boundaries from which an image pixel mask is generated corresponding to all interior points of any boundary. This is computed via local cellular-automata-type spreading rules, with the spread broken across the fences. This allows for the texture feature accumulation.

The masking and fencing are done for each simple move. For the birth moves, since a single object is added to the identical configuration, the old mask is saved with sequential addition of one object at a time, accumulating the new additional pseudolikelihood along with the penalty term associated with overlaps between the hypothesized object and the existing mask. The death moves simply remove one of the existing objects. Again, the masking power of the SIMD machine is ideal.

5.2. *Experimental Results*

Fig. 3 shows varying points during the random sampling algorithm for one of the data sets studied, proceeding in time sequence from left to right. Fig. 4 illustrates the jump moves showing a birth (top row), a death (middle row) and a merger (bottom row). The left-hand panel in each row shows the state of the Markov process at the instant before the jump, with the right-hand panel showing the new configuration after the jump has occurred. Fig. 5 shows the state of the random sampling algorithm after 1000 iterations for all three data sets.

Analysis of the results demonstrates numerous strengths, as well as shortcomings, of the approach. First, the shape models and priors are sufficiently flexible to accommodate the variation in organelle shapes. Secondly the merge, add and delete

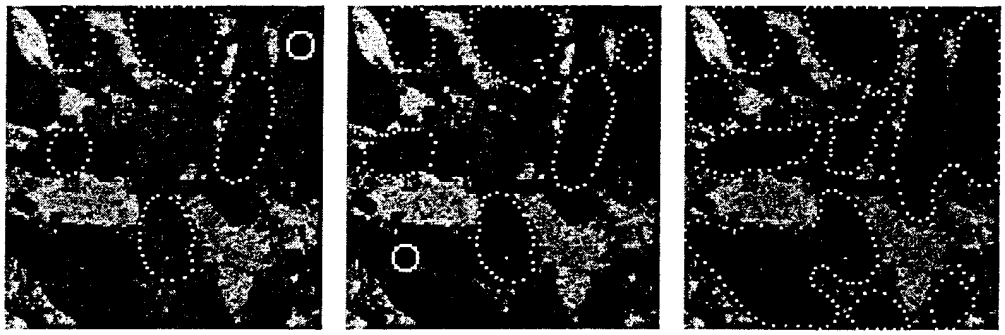


Fig. 3. Various states of the Markov process with time progressing from left to right

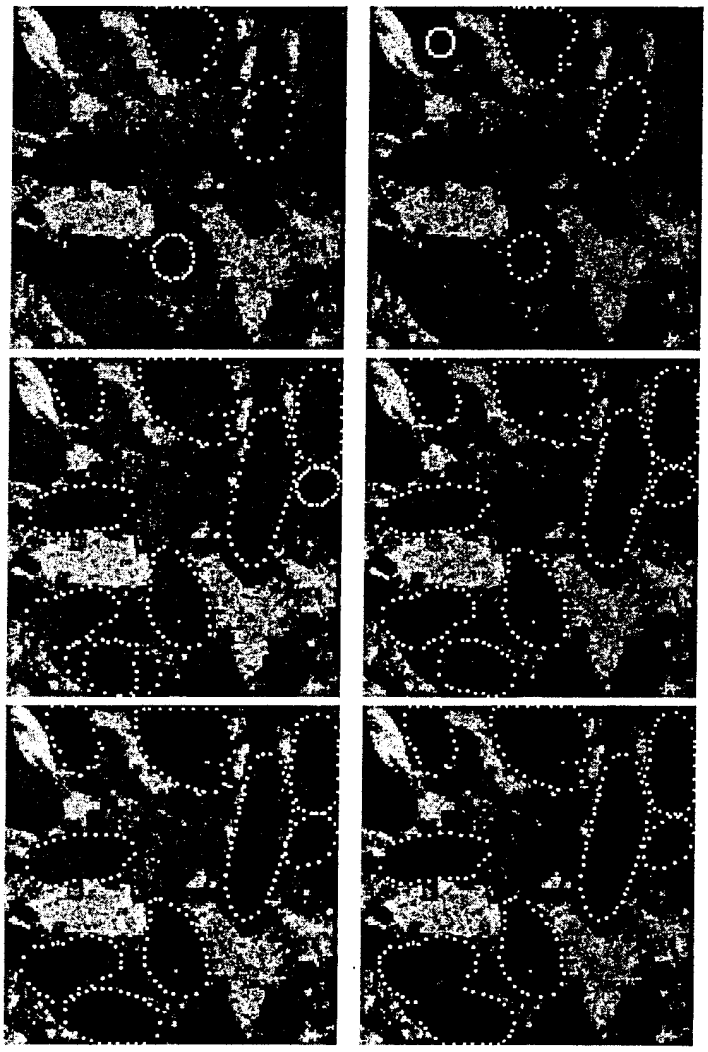


Fig. 4. Birth of mitochondria (top row), death of mitochondria (middle row) and a merging of two mitochondria (bottom row)

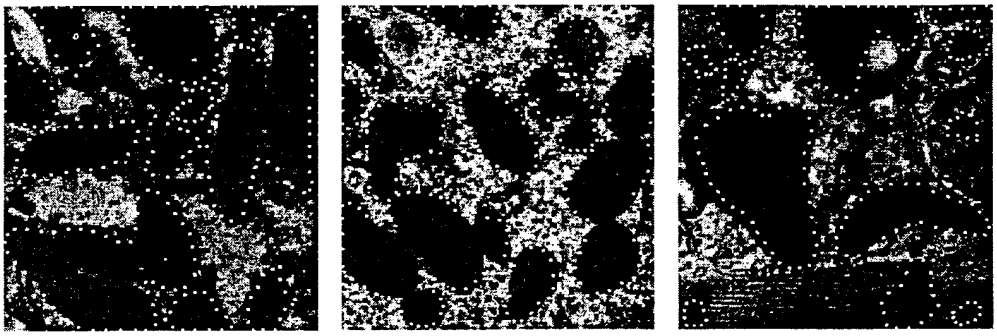
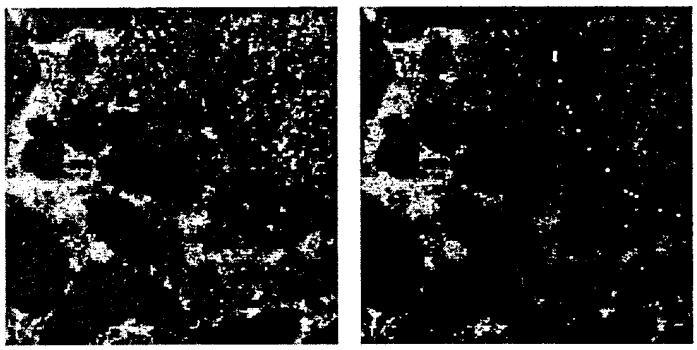


Fig. 5. State of the random sampling algorithm after 1000 iterations of the Markov process for the three data sets in Fig. 1



(a) (b)

Fig. 6. Data containing a linear membrane, with the result of the segmentation using the Markov process and the linear membrane representation

moves are adequate for discovering the correct model order for the scenes. A clear inadequacy of the method is the application of only two texture types, mitochondria and background cytoplasm: other organelles are clearly evident. The centre organelle in the left-hand panel of Fig. 5 is rough endoplasmic reticulum. Because of its high concentration of proteins it is mistaken for mitochondria. The algorithm does not at present distinguish between the two. Fig. 6 shows the result of using the linear graph model for segmenting membrane organelles. Fig. 6(a) shows the micrograph data, with Fig. 6(b) showing the result of running the Markov process constraining the model to contain a single linear structure of arbitrary growing length. In running the membrane program, the algorithm was seeded with a point in the scene from which the membrane is grown.

A second problem with the algorithm is its tendency to join abutting objects. Notice the clearly delineable mitochondria which have been merged in all the segmented data in Fig. 5. On careful consideration it appears that humans distinguish close objects via the long narrow gulfs which form between the objects. A fourth graph change, a *cut–move* is being implemented by introducing linear membrane structures which act as scissors for cutting adjoining objects.

6. DISCUSSION

The pattern theoretic methods are being applied in several applications: HANDS (Grenander *et al.*, 1990), LEAVES (Knoerr, 1988), XRAYS (Amit *et al.*, 1991), NMR (Miller and Greene, 1989; Chen *et al.*, 1993), AMOEBA (Joshi *et al.*, 1992; Joshi and Miller, 1993), BRAINS (Grenander *et al.*, 1992; Christensen *et al.*, 1993a, b; Miller, Christensen, Amit and Grenander, 1993), TRACKS (Miller and Fuhrmann, 1990; Srivastava *et al.*, 1991, 1992; Miller, Teichman, Srivastava, O'Sullivan and Snyder, 1993), LANGUAGE (Mark and Miller, 1992), ARTERIES (Elion *et al.*, 1991) and COLOR (Grenander and Manbeck, 1993). We shall briefly describe several applications in HANDS, LEAVES and XRAYS, which are projects that have been virtually completed, and then show results from problems in BRAINS and TRACKING, both of which feature the global shape model and random graph (unknown model order) estimation components of the problem. This emphasizes the context in which the jump-diffusion random sampler would find application.

In HANDS human hands were captured by a digital camera with the boundary generators directed line segments in R^2 with two bond values their end points. The transformations were from the translation and rotation groups similar to the mitochondria problem. The connector graph $\sigma = n$(cyclic) was fixed with n sites. The prior was Gaussian, invariant with respect to cyclic permutations of the n sites, and specialized to be a cyclic Markov process. The inference was achieved by using stochastic relaxation. A related example, LEAVES, deals with the shapes of leaves.

In XRAYS the objects were again human hands but pictures were applied by X-ray photography so that the hands appeared as partially transparent. One set of generators was obtained by discretizing the unit square and using the lattice points as generators. On these operated various linear groups, predominantly translation groups. The translation groups applied to all lattice points were parametrically constrained to lie in a basis formed by normalized eigenfunctions of the discrete Laplacian, with the inference implemented through Langevin diffusion.

6.1. *BRAINS—Deformable Anatomical Text-books*

We now explore the application of global shape models to the representation of the highly complex systems of neuroanatomies. To illustrate human neuroanatomical variability Fig. 7 shows two T2-weighted MR axial brain slice images from two patients (Figs 7(a) and 7(b)) collected in the Department of Radiology at Duke University. Both images contain the same global structures—white matter, grey matter and ventricles—but differ in overall size and orientation. The shapes of the internal structures also differ. Notice that the four ventricles in Fig. 7(a) are smaller than those of Fig. 7(b) and notice the variations in the folds of grey matter.

In most of the shape representation work already described templates of low complexity have been used which can be constructed with modest effort: *constructing templates for human anatomies is a task that is orders of magnitude bigger.* Until recently the construction of the template itself seemed to be the major obstacle for the successful application of these methods. It was therefore a welcome surprise to learn of the 'Visible human' project (US National Library of Medicine Board of Regents, 1987) undertaken by the National Medical Library (NML) in which digital anatomical templates are being constructed for two complete human beings. Quoting from the NML:

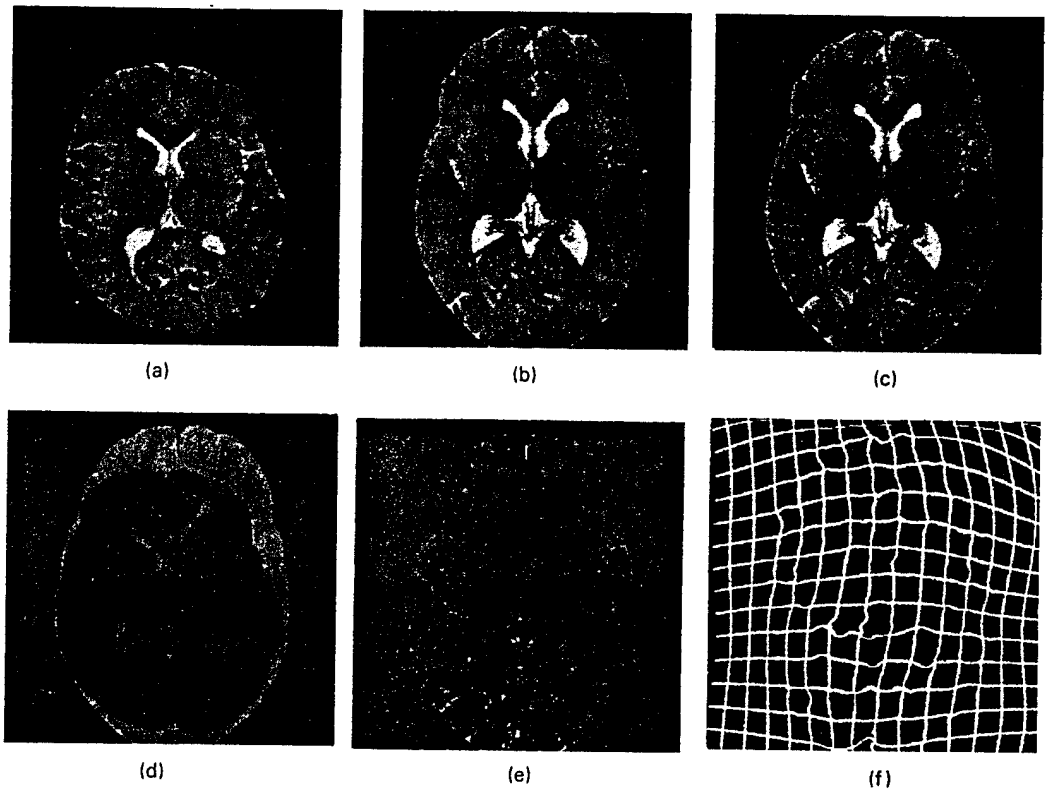


Fig. 7. T2-weighted MR axial brain slice images from (a) the text-book and (b) the patient; (c) result of transforming the text-book co-ordinates into the patient; difference between the text-book and patient (d) before and (e) after transformation of the co-ordinate systems; (f) transformation applied to the text-book co-ordinate system, a rectangular grid

'This Visible Human project would include digital images derived from computerized tomography, magnetic resonance imagery, and photographic images from cryosectioning of cadavers'.

These digital libraries form the basis for the templates of our global shape models as follows.

The anatomical template becomes a multivalued vector function $\mathbf{T}$ defined on the *ideal co-ordinate system* $\Omega \subset R^3$ of the *text-book*: $\mathbf{T}: \Omega \to \mathscr{T}$, with $\mathscr{T}$ the range space assumed to be an M-fold product of spaces $\mathscr{T}_1 \times \mathscr{T}_2 \times \ldots \mathscr{T}_M$, where each component $T_m \in \mathscr{T}_m$ corresponds to a different feature of the tissue. The vector function contains intensity values of various sensor probes: MR spin density, t1 and t2 images, and CT attenuation density. It can also contain physiological and histological information, as well as symbolic information associated with the various labelled areas: white matter tracts, grey matter nuclei, Broca's areas, etc. The triple $(\Omega, \mathbf{T}, \mathscr{T})$ is termed the *anatomical text-book (template)*.

Normal human variation is accommodated by defining a set of transformations generated from translation groups applied to all lattice points in Ω:

$$(x_1, x_2, x_3) \to (x_1 - u_1(\vec{x}), x_2 - u_2(\vec{x}), x_3 - u_3(\vec{x})). \tag{35}$$

The maps constructed from these high dimensional transformations allow for the local mapping of the underlying ideal co-ordinates of the template.

The anatomical text-book is applied to individual patients by assuming that a *patient* is characterized via a *study* **S**, an N-valued vector function consisting of N characterizing data sets $\{S_n\}_{n=1}^N$, or *substudies*. It is assumed that all the study types already exist in the *ideal* text-book: $S_n: \Omega \to \mathscr{T}_{m_n}$, for some $m_n \in \{1, 2, \ldots, M\}$. The information in the anatomical text-book $(\Omega, \mathbf{T}, \mathscr{T})$ is brought into the co-ordinates of the patient by finding the transformation *registering* the studies $\{S_n\}_{n=1}^N$ with the text-book.

Registration between **T** and **S** is defined by using a distance measure evaluated on the transformed text-book and the study. For all the MR data the squared error distance is used which is consistent with Gaussian models of noise in MR imaging. A prior is induced on the transformations to obey linear elasticity for deformable solids. The transformations are generated by using Langevin diffusion on the discrete posterior induced by the potential

$$\sum_{n=1}^{N} \int_{\Omega} |T_n\{\vec{x} - \vec{u}(\vec{x})\} - S_n(\vec{x})|^2 \, d\vec{x} + E(\vec{u}) \tag{36}$$

with

$$E(\vec{u}) = \sum_{i=1}^{3} \sum_{j=1}^{3} \int_{\Omega} \lambda \left\{ \frac{\partial u_i(\vec{x})}{\partial x_i} \right\} \left\{ \frac{\partial u_j(\vec{x})}{\partial x_j} \right\} + \mu \left\{ \frac{\partial u_i(\vec{x})}{\partial x_j} + \frac{\partial u_j(\vec{x})}{\partial x_i} \right\}^2 d\vec{x}. \tag{37}$$

Linear elasticity, isotropy, conservation of momentum and small deformation assumptions yield the energy function (37). See Grenander *et al.* (1992), Miller, Christensen, Amit and Grenander (1993) and Christensen *et al.* (1993a, b) for details.

In our study we take the patient corresponding to Fig. 7(a) to be the text-book. Shown in Fig. 7(c) is the result of transforming the co-ordinate system of the text-book to bring the MR modalities between the two patients into register. Notice the close correspondence between the patient (Fig. 7(b)) and the transformed text-book (Fig. 7(c)). To illustrate the closeness of the two brain slices, shown in Figs 7(d)–7(f) are difference images between the patient and text-book before (Fig. 7(d)) and after (Fig. 7(e)) transformation of the co-ordinate systems. Fig. 7(f) shows the transformation applied to the text-book co-ordinate system, a rectangular grid. Notice how both the global as well as local variation has been accommodated by the transformation.

6.2. *TRACKS: Automated Tracking Target Recognition*

A fundamental task in the representation of complex dynamically changing scenes involving rigid targets is the construction of models that accommodate the variability of orientation, range, object number and object type. The problem is to track and identify the orientation, translation and scale parameters accommodating the variability manifest in the viewing of each object type. See Srivastava *et al.* (1991, 1992) and Miller, Teichman, Srivastava, O'Sullivan and Snyder (1993) for details.

The second distinct part of the sampling process is associated with choosing the target types. The deduction algorithm must go through multiple stages of hypothesis during which the airplane types are being discovered. This is accommodated by using the jump transformations from one object type to another, where a jump may

correspond to the hypothesis of a new object in the scene, or a 'change of mind' about an object type.

The subset of generators $\mathscr{G}^{(0)} \subset \mathscr{G}$ are the templates, airplane surfaces defined by a surface lattice indicating the position in three-dimensional space of every point and its normal pointing direction. Shown in Fig. 8(a) is one such generator from $\mathscr{G}^{(0)}$. These are transformed via the application of operators $s(\vec{p})$, $\vec{p} \in R^3$, in the translation group and $s(\vec{\phi})$ in the orthogonal group parameterized by $\vec{\phi} \in [0, 2\pi]^3$ pitch, roll and yaw in the toral group, 0, 2π identified.

The parameterized transformations operate globally on the template targets of $\mathscr{G}$ generating the full target space. The fact that these transformations involve the *orthogonal* group which we parameterize through the *torus* emphasizes the need for diffusions in spaces more general than Euclidean spaces as provided by our theorem 1. These results have been generalized to Lie manifolds in Amit *et al.* (1993). The parameters specify similarity transformations, as well as the airplane type, with each generator $g \in \mathscr{G}$ specified via a parameter vector $x \in [0, 2\pi]^3 \times R^3 \times \mathscr{A}$, $\mathscr{A}$ the alphabet of target types.

We are interested in tracking and recognition in 'hostile or non-co-operative' environments in which the objects can appear and disappear at random times, implying that tracks will be over varying length. The group transformations are made finite

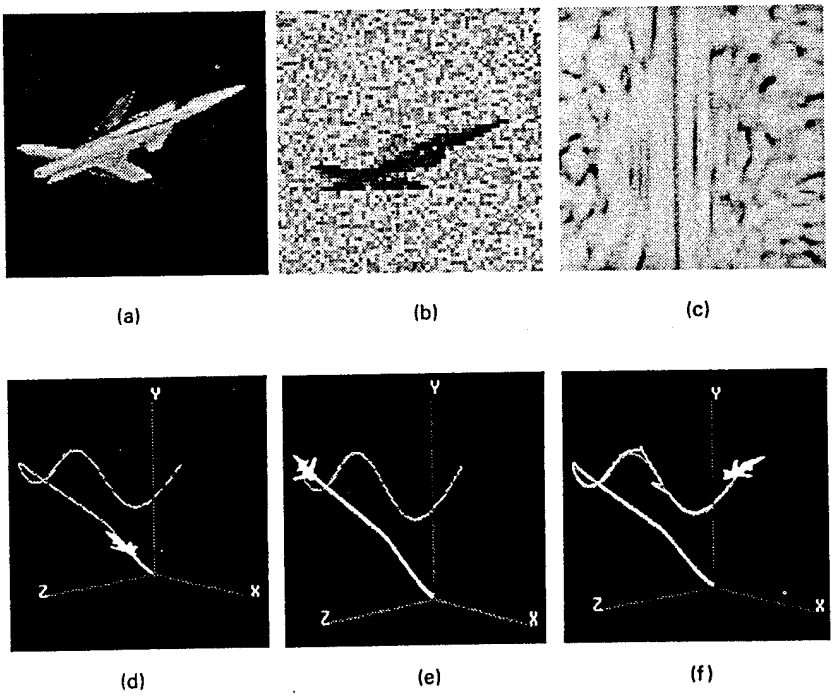


Fig. 8. (a) Three-dimensional single-target generator at a fixed point in time; (b) high resolution optical data; (c) azimuth–elevation signal power profile generated from the narrow band tracking data; (d) actual track with the estimated target superimposed; (e), (f) successive stages of algorithm position, orientation and track identification

dimensional via association with a discrete set of times giving the parameter vector associated with an n-length track $x(n)\in\mathscr{G}(n)\equiv([0,\ 2\pi)^3\times R^3\times\mathscr{A})^n$. Since n is unknown, the full parameter space becomes $\mathscr{G}\equiv\cup_{n=0}^{\infty}([0,\ 2\pi)^3\times R^3\times\mathscr{A})^n$. In the problem stated here the data $I^{\mathscr{D}}$ have multiple components corresponding to the various sensors: $I^{\mathscr{D}}$ refers to narrow band tracking and $I_2^{\mathscr{D}}$ to high resolution optics.

6.2.1. *Tracking priors*

The prior on tracks is based on the dynamics of target motion and follows that described in Srivastava *et al.* (1992) in which the force equations governing the motion of targets are utilized to form a prior density on the track parameter space. The prior on Newtonian dynamics is in terms of the velocity of the airplane $\vec{v}(t)\in R^3$ and is related to the translation group parameters according to

$$\vec{p}(t)=\int_{t_0}^{t}\psi(\tau)\,\vec{v}(\tau)\,d\tau+\vec{p}(t_0),$$

where $\psi(\tau)$ is the 3×3 rotation matrix. Under simplifying assumptions (Friedland, 1986; Srivastava *et al.*, 1992) (rigid body, earth's curvature and wind negligible) the velocity satisfies

$$\dot{\vec{v}}(t)+A\{\vec{\phi}(t),\ \dot{\vec{\phi}}(t)\}\vec{v}(t)=\vec{f}(t) \tag{38}$$

where $\vec{f}$ is the forcing function. The set of Euler angles $\vec{\phi}(t)\in[0,\ 2\pi]^3$ represents the orientation of the target with respect to its body frame,

$$A\{\vec{\phi}(t),\ \dot{\vec{\phi}}(t)\}=\begin{pmatrix} 0 & -q_3(t) & q_2(t) \\ q_3(t) & 0 & -q_1(t) \\ -q_2(t) & q_1(t) & 0 \end{pmatrix},$$

and $\vec{q}(t)$ are the angular velocities in the body frame, functions of the Euler angles.

The covariance is induced following the approach in Amit *et al.* (1991) and Srivastava *et al.* (1992) by assuming that the forcing function is white inducing a Gaussian process $\vec{v}(t)$ with covariance operator determined by the differential operator of equation (38). Since the time varying parameter matrix $A\{\vec{\phi}(t),\ \dot{\vec{\phi}}(t)\}$ is parameterized by the sequence of airplane orientations $\vec{\phi}(t)$, $t\in[0,\ T)$, the tracking and recognition algorithms are linked.

6.2.2. *The likelihood: tracking and imaging data*

There are two sensor types: a *tracking* and a *high resolution imaging* sensor. For tracking, a narrow band array as in Miller and Fuhrmann (1990) and Srivastava *et al.* (1991, 1992) is assumed using the standard narrow band signal model developed in Schmidt (1981). The cross consists of two uniform, linear orthogonal arrays, sensitive to the range, elevation and azimuth locations of the targets. The data collected at the P-element sensor array at time t become the superposition of the incoming signal and the ambient noise. The *deterministic signal model* is used (Miller and Fuhrmann, 1990; Srivastava *et al.*, 1991, 1992) in which the measurements $\mathbf{y}(t)$ are Gaussian

distributed with mean determined by the Vandermonde direction vector associated with the linear arrays and parameterized by the azimuth and elevation angles.

We are currently incorporating models for high resolution radar imaging as described in Snyder *et al.* (1989), Miller *et al.* (1990) and Moulin *et al.* (1992). All the results shown are based on optical imaging systems in which the data are a two-dimensional projection with additive noise. Figs 8(b) and 8(c) show the data at one time instant. Fig. 8(b) shows the high resolution optical images; Fig. 8(c) shows the spatial power spectrum from the narrow band tracker plotted in the azimuth–elevation plane (bright is low power, dark is high power).

The tracking recognition algorithm was implemented by using a Silicon Graphics workstation for data generation and visualization, and the 4096 processor SIMD DECmpp 12000 SX Model 200 machine for implementing the tracking recognition random sampling algorithm. A narrow band cross consisting of two 32-element arrays was simulated with high resolution optical data generated at every time instant. The track estimation proceeds by births and deaths of track segments at random times, with the stochastic gradient search running between the jumps for adjusting the orientation and position estimates. Figs 8(d)–8(f) display successive stages of the algorithm, with the estimated track superimposed over the actual track in white along with the estimated target type and orientation.

7. CONCLUSION

Basing our approach on pattern theoretic representations of image ensembles, in particular global shape models, we have built an algorithm that carries out automatic hypothesis formation for organelles in electron micrographs. The algorithm gives one or several explanations of the picture in terms of the number of mitochondria and membranes, their locations, orientations and shapes. The understanding realized by the algorithm can be claimed to be successful in that in doubtful (for the algorithm) cases a human observer also hesitates between alternative explanations. The algorithm could be made to give degrees of belief for the alternative explanations.

This, together with recent work on neuroanatomies tracking and target recognition, leads us to believe that the general ideas of this approach are applicable widely, and also to non-pictorial patterns. As we proceed to increasingly complex image ensembles the primary task is to build pattern theoretic representations of the ensembles. If this can be done, incorporating subject-matter knowledge in a precise and realistic way, the analytical and computational issues that will arise can be dealt with. We therefore have a powerful and practical technology for pattern analysis in a variety of applications.

ACKNOWLEDGEMENTS

We would like to express our special appreciation to Yali Amit of the University of Chicago who assisted on various technical aspects of the convergence of the random sampling algorithm.

We would like to express our appreciation to Jay Shrauner and Dave Maffitt who played key roles in the parallel implementation on the SIMD processor, to Daniel Fuhrmann for suggestions concerning block circulant diagonalizations, to Scott Nadel of Duke University for providing the MR data sets and to Antonio Possolo of the Boeing Corporation for many helpful suggestions during the preparation of the

manuscript. This research was supported by grants ARO DAAL-03-86-K-0110, ONR N00014-88-K-289, NIH-NCRR-RR01380, ARO DAA-03-92-G-0141 and ONR N00014-92-J-1418, and Rome Laboratory grant F30602-92-C-004.

REFERENCES

Amit, Y., Grenander, U. and Miller, M. I. (1993) Ergodic properties of jump-diffusion processes. *Technical Report 361*. Department of Statistics, University of Chicago, Chicago.

Amit, Y., Grenander, U. and Piccioni, M. (1991) Structural image restoration through deformable templates. *J. Am. Statist. Ass.*, **86**, 376–387.

Apostol, T. M. (1974) *Mathematical Analysis*, p. 356. Reading: Addison-Wesley.

Athreya, K. B. and Ney, P. (1978) A new approach to the limit theory of recurrent Markov chains. *Trans. Am. Math. Soc.*, **245**, 493–501.

Bajcsy, R. and Kovacic, S. (1989) Multiresolution elastic matching. *Comput. Vis. Graph. Im. Process.*, **46**, 1–21.

Bartlett, M. S. (1975) *The Statistical Analysis of Spatial Patterns*. London: Chapman and Hall.

Besag, J. (1974) Spatial interaction and statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.

Besag, J. and Green, P. J. (1993) Spatial statistics and Bayesian computation. *J. R. Statist. Soc.* B, **55**, 25–37.

Bookstein, F. L. (1978) The measurement of biological shape and shape change. *Lect. Notes Biomath.*, **24**.

Chen, S. C., Schaewe, T. J., Teichman, R. S. and Miller, M. I. (1993) Parallel algorithms for maximum likelihood nuclear magnetic resonance spectroscopy. *J. Magn. Reson.* A, **102**, 16–23.

Chomsky, N. (1956) Three models for the description of language. *IRE Trans. Inform. Theory*, **2**, 113–124.

——(1959) On certain formal properties of grammars. *Inform. Control*, **2**, 137–167.

Christensen, G. E., Rabbitt, R. D. and Miller, M. I. (1993a) 3D brain mapping using a deformable neuroanatomy. *Phys. Med. Biol.*, to be published.

——(1993b) A deformable neuroanatomy textbook based on viscous fluid mechanics. In *Proc. Conf. Information Sciences and Systems*, pp. 211–216. Baltimore: Johns Hopkins University Press.

Elion, J. L., Geman, S. A. and Manbeck, K. M. (1991) Computer recognition of coronary arteries. *J. Am. Coll. Card.*, **17**, no. 2.

Ethier, S. N. and Kurtz, T. G. (1986) *Markov Processes*. New York: Wiley.

Feller, W. (1936) Zur Theorie der stochastischen Prozesse. *Math. Ann.*, **113**, 113–160.

Fosdick, L. D. (1963) *Monte Carlo Computations on the Ising Model*. New York: Academic Press.

Freiberger, W. and Grenander, U. (1969) Computer generated image algebras. *Int. Fed. Inform. Process.*, **68**, 1397–1404.

Friedland, B. (1986) *Control System Design: an Introduction to State-space Methods*. New York: McGraw-Hill.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to computing marginal densities. *J. Am. Statist. Ass.*, **85**, 398–409.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.

Geman, S. and Hwang, C.-R. (1987) Diffusions for global optimization. *SIAM J. Control Optimizn*, **24**, 1031–1043.

Gidas, B. (1993) Metropolis type monte-carlo simulation algorithms and simulated annealing. *Trends Contemp. Probab.*, to be published.

Gihman, I. I. and Skorohod, A. V. (1965) *Introduction to the Theory of Random Processes*. Philadelphia: Saunders.

Goodman, N. R. (1963) Statistical analysis based on a certain multivariate complex Gaussian distribution. *Ann. Math. Statist.*, **34**, 152–177.

Gregory, J. and McReynolds, R. (1963) The SOLOMAN computer. *IEEE Trans. Elect. Comput.*, **12**, 774–781.

Grenander, U. (1967) Probability measures for context-free languages. Brown University, Providence.

——(1970) *A Unified Approach to Pattern Analysis*, vol. 10.

——(1981) *Regular Structures: Lectures in Pattern Theory*, vol. III. New York: Springer.

——(1983) Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence.
——(1985) Advances in pattern theory: the 1985 Rietz lecture. *Ann. Statist.*, **17**, 1–30.
Grenander, U., Chow, Y. and Keenan, D. (1990) *HANDS: a Pattern Theoretic Study of Biological Shapes*. New York: Springer.
Grenander, U. and Manbeck, K. M. (1993) A stochastic model for defect detection in potatoes. *J. Comput. Graph. Statist.*, to be published.
Grenander, U. and Miller M. I. (1991) Jump–diffusion processes for abduction and recognition of biological shapes. *Monograph*. Electronic Signals and Systems Research Laboratory, Washington University, St Louis.
Grenander, U., Miller, M. I. and Christensen, G. (1992) Deformable anatomical data bases using global shape models: a position paper for the 1992 electronic imaging of the human body workshop. In *Proc. Cooperative Working Group on Whole Body 3-D Electronic Imaging of the Human Body*.
Hajek, B. (1988) Cooling schedules for optimal annealing. *Math. Ops Res.*, **13**, no. 2.
Harris, T. E. (1963) *The Theory of Branching Processes*. Berlin: Springer.
Hastings, W. K. (1970) Monte carlo sampling methods using markov chains, and their applications. *Biometrika*, **57**, 97–109.
Horn, P. S. (1977) The problem of simulating random configurations governed by the gibb's ensemble. *Pattern Analysis Report 49*.
Joshi, S. and Miller, M. I. (1993) Maximum *a posteriori* estimation with Good's roughness for optical sectioning microscopy. *Opt. Soc. Am.* A, **10**, 1078–1085.
Joshi, S., Miller, M. I., McNally, J., Amit, Y. and Grenander, U. (1992) Global shape models for optical sectioning microscopy. *J. Opt. Soc.*, Oct.
Kendall, D. G. (1977) The diffusion of shape. *Adv. Appl. Probab.*, **9**, 428–430.
——(1984) Shape manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81–121.
Knoerr, A. (1988) Global models of natural boundaries. *Pattern Analysis Report 148*.
Mark, K., Miller, M. I., Grenander, U. and Abney, S. (1992) Parameter estimation for constrained context-free language models. *1992 DARPA Wkshp Speech and Natural Language, Feb.*
Mead, C. (1989) *Analog VLSI and Neural Systems*. Reading: Addison-Wesley.
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller A. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Phys. Chem.*, **21**, 1087.
Miller, M. I., Christensen, G. E., Amit, Y. and Grenander, U. (1993) A mathematical textbook of deformable neuro-anatomies. Submitted to *Proc. Natn. Acad. Sci. USA*.
Miller, M. I. and Fuhrmann, D. R. (1990) Maximum likelihood narrow-band direction finding and the EM algorithm. *IEEE Acoust. Spch Signal Process.*, **38**, 560–577.
Miller, M. I., Fuhrmann, D. R., O'Sullivan, J. A. and Snyder, D. L. (1990) Maximum-likelihood methods for toeplitz covariance estimation and radar imaging. In *Advances in Spectrum Estimation* (ed. S. Haykin), pp. 145–172. Englewood Cliffs: Prentice Hall.
Miller, M. I. and Greene, A. (1989) Maximum-likelihood estimation for nuclear magnetic resonance spectroscopy. *J. Magn. Reson.*, **83**, 525–548.
Miller, M. I., Larson, K. B., Saffitz, J. E., Snyder, D. L. and Thomas, Jr, L. J. (1985) Maximum-likelihood estimation applied to electron-microscope autoradiography. *J. Electron Microsc. Tech.*, **2**, 611–636.
Miller, M. I. and O'Sullivan, J. A. (1992) Entropies and combinatorics of random branching processes and context-free languages. *IEEE Trans. Inform. Theory*, **38**, no. 4.
Miller, M. I., Roysam, B., Smith, K. and O'Sullivan, J. A. (1991) Representing and computing regular languages on massively parallel networks. *IEEE Trans. Neural Netwks*, **2**, 56–72.
Miller, M. I., Teichman, R. S., Srivastava, A., O'Sullivan, J. A. and Snyder, D. L. (1993) Jump–diffusion processes for automated tracking-target recognition. In *Proc. Conf. Information Sciences and Systems*, pp. 617–622. Baltimore: Johns Hopkins University Press.
Moulin, P., O'Sullivan, J. A. and Snyder, D. L. (1992) A method of sieves for multiresolution spectrum estimation and radar imaging. *IEEE Trans. Inform. Theory*, to be published.
Netter, F. H. (1980) *The CIBA Collective of Medical Illustrations*. Summit: Ciba.
Revuz, D. (1975) *Markov Chains*. Amsterdam: North-Holland.
Ripley, B. D. (1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
Rissanen, J. (1987) Stochastic complexity (with discussion). *J. R. Statist. Soc.* B, **49**, 223–239, 253–265.

Roysam, B. and Miller, M. I. (1992) Combining stochastic and syntactic processing with analog computation methods. *Dig. Signal Process.*, **2**, 48–64.
Schmidt, R. (1981) A signal subspace approach to multiple emitter location and spectral estimation. *PhD Dissertation*. Stanford University, Palo Alto.
Smith, K. and Miller, M. I. (1989) Learning regular grammars on connection architectures. In *ICASSP-89*, pp. 2501–2504.
——(1990) *Bayesian Inference of Regular Grammar and Markov Source Models*, pp. 388–398. Palo Alto: Morgan Kaufmann.
Snyder, D. L., O'Sullivan, J. A. and Miller, M. I. (1989) The use of maximum-likelihood estimation for forming images of diffuse radar-targets from delay-doppler data. *IEEE Trans. Inform. Theory*, **35**, 536–548.
Srivastava, A., Cutaia, N., Miller, M. I., O'Sullivan, J. A. and Snyder, D. L. (1992) Multi-target narrowband direction finding and tracking based on motion dynamics. In *Proc. 30th A. Allerton Conf. Communication, Control and Computing, Urbana.*
Srivastava, A., Miller, M. I. and Grenander, U. (1991) Jump–diffusion processes for object tracking and direction finding. In *Proc. 29th A. Allerton Conf. Communication, Control and Computing, Urbana*, pp. 563–570.
Staib, L. H. and Duncan, J. S. (1992) Boundary finding with parametrically deformable models. *IEEE Trans. Pattn Anal. Mach. Intell.*, **14**, 1061–1075.
Terzopoulos, D. and Waters, K. (1990) Physically-based facial modelling, analysis, and animation. *J. Vis. Comput. Animn*, **1**, 73–80.
Thompson, d'A. W. (1917) *On Growth and Forms.* Cambridge: Cambrige University Press.
US National Library of Medicine Board of Regents (1987) *National Library of Medicine, Long Range Plan: ELECTRONIC IMAGING.* Bethesda: US Department of Health and Human Services.
Whittle, P. (1954) On stationary processes in the plane. *Biometrika*, **41**, 434–449.
Wright, S. (1921) System of mating. *Genetics*, **6**, 111–178.
Wyatt, J. L. (1992) Analog vlsi systems for early vision. *IEEE Int. Symp. Circuits and Systems.*

DISCUSSION OF THE PAPER BY GRENANDER AND MILLER

John T. Kent (University of Leeds): Ulf Grenander has played a pioneering role in the development and application of probabilistic methods in image analysis and related areas. Indeed he has inspired a whole generation of research workers. Michael Miller combines an appreciation of the importance of careful probabilistic modelling with the practical viewpoint of an engineer. Together they make a powerful team and this paper offers a flexible and sophisticated approach for the recognition of structure in images.

I want to focus my comments on two aspects of the paper. The first point is a comparison of three different ways to represent object outlines in the plane: vertices, edges and similarities. The second point is a study of the parsimonious description of object variability. It is important to keep models as simple as possible, both to facilitate interpretation and to make the extension to new applications as effective as possible.

To begin, consider the outline of an object in R^2 with n vertices $\{v_j, j=1, \ldots, n\}$. The vertices may be physically identifiable landmarks on the object, or they may be n arbitrary equally spaced points around the outline (as for the mitochondria example). Two other ways to represent the outline are in terms of the edges $e_j = v_j - v_{j-1}$ or in terms of the 'similarities' described in Section 2.1.2. For both edges and similarities, it is necessary to add an overall location vector to complete the representation of the outline. The similarities essentially measure the relative change in each edge with respect to an underlying template.

All three representations are linearly related to one another. Therefore, it is largely a matter of taste which representation is used. In many applications the vertex representation is most straightforward. However, a notable case where the similarity representation is most elegant is given by the mitochondria example in the paper, with its underlying rotational symmetry.

Next I would like to turn to the topic of parsimonious models. A general covariance matrix in n dimensions requires $O(n^2)$ pieces of information for large n. Therefore there is a need for methods of expressing variability by using a smaller number of parameters.

The most straightforward approach involves the graph structures of Section 2 so that only neighbouring generators contribute an interaction term to the probability density function. In the context of a Gaussian

model for edges or similarities around an outline, this model reduces to a first-order cyclic Markov random field, as used in the HANDS project of Grenander *et al.* (1990). Unfortunately, the edge process needs to be constrained, $\Sigma e_j = 0$, which destroys its Markov nature. However, in this case, provided that the overall position of the object is given a uniform prior in R^2, the edge model can be recast as a second-order improper cyclic Markov random field model on the vertices (Mardia *et al.*, 1991). Further, for large n, the quadratic form in the exponent of the density is related to geometric quantities of a path such as the integrated squared curvature (Kent *et al.*, 1992).

An alternative route to parsimony is through principal components (e.g. Kent (1994) and Cootes *et al.* (1992)). In many examples the variation in the data is dominated by the first few principal components of the covariance matrix. Further, for large n and with smooth outlines, these principal components are often analogous to low frequency terms in a Fourier series.

As recognized by the authors in their mitochondria example, one attempt to achieve parsimony that is often unsuccessful is the assumption of complex symmetry. This assumption is very attractive from an analytical point of view but can be unrealistic in practice. For example, the amount of variability at a vertex in the tangent direction of the outline may be different from the variability in the normal direction. See also Kent (1994) for further discussion.

Next we turn to the circulant covariance matrix K in Section 4.1.1 for the mitochondria. Here the assumption of rotational symmetry nicely limits the number of parameters. As the authors point out, the underlying structure is brought out most clearly by transforming to the independent four-dimensional vectors $(\mathrm{Re}\{\bar{u}(k)\}, \mathrm{Re}\{\bar{v}(k)\}, \mathrm{Im}\{\bar{u}(k)\}, \mathrm{Im}\{\bar{v}(k)\})$, $k = 0, \ldots, n/2$. These vectors have covariance matrices of the form

$$\begin{pmatrix} A(k) & B(k) \\ -B(k) & A(k) \end{pmatrix},$$

where A is 2×2 symmetric and B is 2×2 skew symmetric (except for $k = 0, n/2$, where the imaginary parts of $\bar{u}(k)$ and $\bar{v}(k)$ vanish).

Since K is a symmetric matrix, the 2×2 blocks satisfy $K(i) = K(n-i)^{\mathrm{T}}$, $1 \leqslant i \leqslant n/2$, and $K(0)$ is symmetric. Further simplification occurs if the additional assumption is made of 'axial symmetry' or 'mirror symmetry', so that $K(i) = K(n-i)$, $1 \leqslant i \leqslant n/2 - 1$, implying $B(k) = 0$ in the above representation. Such an assumption seems natural for objects such as mitochondria.

Another simplification involves the incorporation of Markov random field structure into K: just set most of the 2×2 blocks in a model for K^{-1} equal to 0. Similarly, principal component analysis can be used to simplify a model by assuming that many of the $A(k)$ and $B(k)$ equal 0. Have the authors considered simplifying K in this way?

The authors have given an impressive panoramic view of their vision of image analysis. In particular the jump diffusion processes represent an elegant application of Markov chain Monte Carlo methodology. At the same time the implementation of their ideas in practice involves many detailed considerations that there has not been space to cover here. I encourage the authors to make available a more complete account of their algorithms to enable other researchers to build on their work.

As you can see, I have found the paper very stimulating and it gives me great pleasure to propose the vote of thanks.

P. Clifford (Oxford University): Traditionally in this Society, the role of the seconder is to be critical of the material which has been presented. There is a small complication here, in that I was probably one of the most enthusiastic supporters of the paper at the review stage. At least, this shows that some of the members of the Research Section Committee have a sense of humour.

A seconder usually starts by posing a fairly fundamental question, such as: 'Is this really statistics?'. He or she then goes on to make a damning observation, e.g. 'Physicists have been doing this for years'. The authors are then chastised for omitting to mention some recent relevant work, and with these preliminaries out of the way the seconder will take the opportunity to talk about his or her own work in the area.

So, is this really statistics? This begs the question: 'What is statistics?'. The authors have suggested that statistics is about understanding information and I agree with this to an extent. Certainly, there are data sets, such as the stack loss figures, which statisticians now understand quite well. These are data whose idiosyncracies have been explored, which are familiar and which have the status of old friends. I suspect that this depth of understanding is atypical. Far more frequently, statistics is about acting on information, and, for this, familiarity with the method is more important than familiarity with the

data. I would question whether the authors are providing real understanding of the data rather than the means to act, for example, to identify the cells.

The paper is important because it is a signpost to future directions in statistics. It illustrates the effectiveness of the traditional engineering 'get it done' philosophy. Just as chemical engineers deal with chemicals in bulk, so information engineers are being trained, in engineering and computer science departments, to handle large data sets with high dimensional parameters. Will we as statisticians be superseded by the information engineers or will we become the architects, the visionaries who design the structure, leaving the construction to the engineers? This paper suggests that such a symbiosis is possible and it may be essential to the survival of our specialized discipline. The principles of statistical data analysis have been absorbed by the engineering community; they have the computing skills, both in hardware and software, to build data analysers which can do the job.

I was reminded of this recently when I read that the US Immigration and Naturalization Service is currently evaluating a system (INSPASS) for identifying individuals to reduce the large queues on entering the USA. An individual puts his or her hand on a plate, and after 30 seconds the hand is identified. This is something put together by engineers. It does not use the beautiful and elegant HANDS theory which Professor Grenander has built up—it just works.

Let us wonder whether we are happy with this, and content to have important statistical problems—data problems—handled by engineers and computer scientists, with essentially no input from the statistical community.

Have physicists been using jump diffusions for years? Well, yes they have, and so have chemists, epidemiologists and those working in point process theory. There are many publications on the problem of simulating the positions and velocities of molecules in finite systems. In the grand canonical ensemble the number of molecules is not fixed but has a distribution given implicitly by the potential. Metropolis and Langevin methods which allow the creation and annihilation of molecules are used; in other words, jump diffusions. In mathematical epidemiology, the locations of the infectious animals are modelled and simulated as a spatial birth–death process, a jump diffusion.

What related recent work has there been? Well, there is the work of Gelfand and Mitter (1991). They showed that certain continuous time interpolations of Metropolis and heat bath Markov chains converge weakly on path space to Langevin diffusions. There are important implications here, in that, if you only wish to sample from the equilibrium distribution then there is no particular benefit in using hybrid jump diffusion processes. You may as well use pure Metropolis sampling, using jumps which are sometimes big and sometimes small.

Finally, my own work: the authors' presentation has emphasized elastic deformations; the cell membrane deforms elastically to fit the picture. I know that in other applications the authors have also considered deformations consistent with viscous flow. My interest is in the estimation of crustal velocity fields on the basis of earthquake data. This is work done in collaboration with Philip England at Oxford. India has been moving north for 20 million years, producing the Himalayas, a great mound of material which then flows onwards forming the Tibetan plateau. The motion is rather slow, about the speed of growth of a fingernail. The crust has to accommodate this motion and the occurrence of earthquakes is related to the associated deformation. There is a substantial database of earthquakes, numbering tens of thousands of events. The objective is to use these data to estimate the underlying velocity field, a field which is modelled by viscous flow theory. We are having some success with this project, and even with the computing resources that we have in Britain we can do the calculations in realtime!

Let me add my congratulations to the authors on producing a stimulating paper. It gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

K. V. Mardia (University of Leeds): Ulf Grenander's school has pioneered many ideas and particularly we in Leeds have been inspired by their work. We were also fortunate to have Professor Miller as the key speaker at our annual research workshop in 1993. Indeed, this paper is extremely profound with many new ideas. Image understanding as described in this paper is the key in most of medical imaging. Some ideas go back at least to Galton (1878) who constructed composite portraits from photographs to understand whether there was an average face associated with a particular trait. Averaging, interpolating, differencing and caricaturing are also fundamental to image understanding. For landmark methods, the planar deformation underlies these constructions. Consider say caricaturing. Given $\mathbf{x}_i$ old landmarks and $\mathbf{y}_i$ new landmarks, $i=1, \ldots, n$, the objective is to find a smooth transformation

$y_1=\Phi_1(\mathbf{x})$, $y_2=\Phi_2(\mathbf{x})$, which takes old landmarks to new, namely, $\mathbf{x}=(x_1, x_2)\rightarrow\mathbf{y}=(y_1, y_2)$. There are situations where we need to use derivative information at landmarks as shown by Bookstein and Green (1993). Instead of using higher derivatives indirectly as in Bookstein and Green (1993), we could proceed in the following way for each of Φ_1 and Φ_2.

Suppose that $\Phi(\mathbf{x})$ is the kriging predictor, $\mathbf{x}\in\mathbf{R}^2$ with constraints $\Phi(\mathbf{x}_i)=y_{i00}$, $\partial\Phi(\mathbf{x}_i)/\partial x_{1i}=y_{i10}$, $\partial\Phi(\mathbf{x}_i)/\partial x_{2i}=y_{i01}$, $i=1, \ldots, n$, $\mathbf{x}_i=(x_{1i}, x_{2i})$ and based on the generalized covariance function $\sigma(\mathbf{h})$. For $\mathbf{x}=(x_1, x_2)$, let $\partial^{(a+b)}\sigma(\mathbf{x})/\partial x_1^a\partial x_2^b=\sigma^{(a,b)}(\mathbf{x})$. It can be shown that $\Phi(\mathbf{x})$ can be written as (Mardia *et al.*, 1993)

$$\Phi(\mathbf{x})=\beta_0+\beta_1x_1+\beta_2x_2+\sum_r\sum_{j=1}^{n}\mathbf{w}_{jr}\,\sigma^{(r)}(\mathbf{x}-\mathbf{x}_j),$$

where the first summation is over $r=(0, 0)$, $(1, 0)$, $(0, 1)$, and

$$\begin{pmatrix}\mathbf{w}\\ \beta\end{pmatrix}=\begin{pmatrix}\Sigma & \mathbf{Q}\\ \mathbf{Q}^{\mathrm{T}} & \mathbf{0}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{y}\\ \mathbf{0}\end{pmatrix},\quad \Sigma=\begin{pmatrix}\Sigma_{00}, \Sigma_{10}, \Sigma_{01}\\ \Sigma_{10}, \Sigma_{20}, \Sigma_{11}\\ \Sigma_{01}, \Sigma_{11}, \Sigma_{02}\end{pmatrix},\quad \mathbf{Q}=\begin{pmatrix}\mathbf{1}, \mathbf{X}_1, \mathbf{X}_2\\ \mathbf{0}, \mathbf{1}, \mathbf{0}\\ \mathbf{0}, \mathbf{0}, \mathbf{1}\end{pmatrix},$$

with $\mathbf{w}^{\mathrm{T}}=(\mathbf{w}_{00}^{\mathrm{T}}, \mathbf{w}_{10}^{\mathrm{T}}, \mathbf{w}_{01}^{\mathrm{T}})$, $\mathbf{w}_r^{\mathrm{T}}=(w_{1r}, \ldots, w_{nr})$, $\beta=(\beta_0, \beta_1, \beta_2)^{\mathrm{T}}$, $\Sigma_{a,b}=(\sigma^{(a,b)}(\mathbf{x}_i-\mathbf{x}_j))$, $\mathbf{1}^{\mathrm{T}}=(1, \ldots, 1)$, $\mathbf{0}^{\mathrm{T}}=(0, \ldots, 0)$, $\mathbf{y}^{\mathrm{T}}=(\mathbf{y}_{00}^{\mathrm{T}}, \mathbf{y}_{10}^{\mathrm{T}}, \mathbf{y}_{01}^{\mathrm{T}})$, $\mathbf{y}_r^{\mathrm{T}}=(y_{1r}, \ldots, y_{nr})$ and $\mathbf{X}_k^{\mathrm{T}}=(x_{k1}, \ldots, x_{kn})$, $k=1, 2$. Hence, $\Phi(\mathbf{x})$ can now be evaluated. We could select for example $\sigma(\mathbf{h})=|\mathbf{h}|^{2\alpha}$, $1<\alpha<2$, so that the kriging equations are well defined. Note that the limiting case as $\alpha\rightarrow 1$, $\sigma(\mathbf{h})=|\mathbf{h}|^2\log|\mathbf{h}|$, is not strictly valid as the second derivative is discontinuous at $\mathbf{h}=\mathbf{0}$ but, for an approximate approach, see Bookstein and Green (1993) who have inspired our work. My colleague, Dr Rabe, will show in her contribution the fitting of our predictor to an average face profile but for further details see Mardia *et al.* (1993).

I have a few simple questions. I appreciate various properties of jump–diffusion but do we need it when there are about 15 or more non-overlapping objects? Is the systematic search too naïve or expensive in such cases? Does the method work well with a very large number of objects? Given a collection of contours of objects where the landmarks are difficult to pick out, what is the most efficient way (Section 2.3, point (d)) to treat landmark misidentification in estimating a template?

A. J. Baddeley (Centre for Mathematics and Computer Science, Amsterdam, and University of Leiden): The 'random configurations' studied in this paper are *point processes* of geometrical objects, which have been studied extensively in stochastic geometry (Stoyan *et al.*, 1987) and spatial statistics (Ripley, 1981, 1988). Indeed Ripley and co-workers (Molina and Ripley, 1989; Ripley, 1986, 1991; Ripley and Sutherland, 1990) have applied point process methods to computer vision, and I was surprised to see no reference to this. Miss M. N. M. van Lieshout and I have also studied object recognition from this viewpoint (Baddeley and van Lieshout, 1991, 1992a, b, 1993; van Lieshout, 1991, 1993) and have some results that are complementary to those of the present paper.

The point process analogue of a Markov random field is a Markov point process (Ripley and Kelly, 1977; Ripley, 1988, 1989). The simplest form is a pairwise interaction process, where a configuration $\mathbf{x}=\{x_1, \ldots, x_n\}$ has density

$$p(\mathbf{x})\propto\prod_{x_i\sim x_j}a(x_i, x_j)$$

where $x_i\sim x_j$ signifies, say, that x_i and x_j are close, $\|x_i-x_j\|<r$. This resembles the authors' equation (5), but Markov point processes may also have higher order interaction terms $a(x_i, x_j, x_k)$ etc., and dynamic graphical interaction structures (Baddeley and Møller, 1989).

A Markov point process can be simulated by running a *spatial birth-and-death process* to equilibrium (Preston, 1977; Ripley, 1977; Møller, 1989). This is a jump process in the space of configurations $\mathbf{x}$, involving instantaneous 'births' (addition of a new point at a random position) and 'deaths' (deletion of an existing point). Other methods are possible, including Metropolis–Hastings algorithms and jump–diffusion processes.

The object recognition problem is to estimate the true configuration $\mathbf{x}=\{x_1, \ldots, x_n\}$, usually a pattern of objects in *continuous* space, from the data $\mathbf{y}$, usually a *discrete* pixel image. Postulating a suitable 'noise model' for $\mathbf{y}$ given $\mathbf{x}$ with density $f(\mathbf{y}|\mathbf{x})$, and a prior distribution $p(\mathbf{x})$ which is a Markov point process, the posterior

$$p(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})$$

is again a Markov point process.

To determine the maximum *a posteriori* estimator of $\mathbf{x}$ we recursively update the current estimate $\mathbf{x}$ by considering the posterior likelihood ratio

$$\frac{p(\mathbf{x}\cup\{a\}|\mathbf{y})}{p(\mathbf{x}|\mathbf{y})} = \frac{p(\mathbf{x}\cup\{a\})}{p(\mathbf{x})} \frac{f(\mathbf{y}|\mathbf{x}\cup\{a\})}{f(\mathbf{y}|\mathbf{x})},$$

updating either deterministically (an algorithm analogous to Besag's iterated conditional modes) or stochastically: the natural analogue of stochastic annealing is a spatial birth-and-death process with transition rates depending on the posterior likelihood ratios. We have studied these algorithms in the papers cited above.

This approach also provides an appreciation of some existing techniques in computer vision. We have shown that, for simple models of binary noise, the maximum likelihood estimator $\hat{\mathbf{x}}$ coincides with the standard erosion and dilation operators of mathematical morphology (Serra, 1982), and, for models such as Gaussian additive noise, the log-likelihood ratio

$$\log f(\mathbf{y}|\mathbf{x}\cup\{u\}) - \log f(\mathbf{y}|\mathbf{x})$$

is the celebrated *Hough transform*, used in computer vision to detect simple features such as lines and circular arcs (Illingworth and Kittler, 1988).

M. N. M. van Lieshout (Free University Amsterdam and Centre for Mathematics and Computer Science, Amsterdam): In addition to Professor Baddeley's contribution I would like to mention a few results from our research.

To find an approximate maximum *a posteriori* (MAP) solution, we proposed to search iteratively for that object whose addition or deletion would most increase the posterior likelihood ratio and to update the scene accordingly (Baddeley and van Lieshout, 1992a, b). This is a variant of Besag's iterated conditional modes (ICM) algorithm for discretized images but is also defined on more general spaces.

The authors propose a jump–diffusion process to sample from the posterior distribution. The jump part bears a close resemblance to the spatial birth-and-death processes introduced by Preston (1977). These are well known in spatial statistics (Ripley, 1977) for sampling Markov point processes. The rate of convergence was studied by Møller (1989). Alternatively, a Metropolis–Hastings algorithm (Geyer and Møller, 1993) is built as a mixture of two transition kernels: one can be regarded as the analogue of Grenander and Miller's diffusion process; the other generates new hypotheses.

The main advantage of sampling from the posterior distribution is the ability to estimate any functional of the posterior (see Section 1.2). In particular, the (estimated) first-order intensity surface can be regarded as an alternative to the Hough transform.

A sequence of birth-and-death processes can be combined in a stochastic annealing schedule. For $H>0$ define

$$p_H(\mathbf{x}|\mathbf{y}) \propto \{f(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})\}^{1/H}.$$

As for discrete Markov random fields, H has the interpretation of 'temperature'. If the set of MAP solutions has positive reference measure and the number of objects is effectively bounded above, a sequence can be constructed that converges in total variation to a uniform distribution on the set of global maxima of the posterior distribution, regardless of the initial state (van Lieshout, 1994).

When H is very close to 0, the corresponding birth–death process behaves like the deterministic algorithm described above. This suggests using an algorithm which incorporates a search operation. However, there will be problems with the 'curse of dimension': as the dimension of the object space increases, the cost of searching it increases exponentially. To overcome this problem, we propose a multiresolution strategy (Baddeley and van Lieshout, 1993).

We have implemented the method and found that it performs creditably on simple test examples. The introduction of a Markov prior successfully combats the multiple-response problem and increases robustness to noise and initial scene selection. For digitized images, replacing the ICM pixel scan by steepest ascent also improves robustness and frees the technique of scanning order dependence. Moreover, Markov point processes are well suited to the techniques proposed, as posterior ratios are typically easy to evaluate.

Ian L. Dryden (University of Leeds): How do the authors obtain estimates of the prior model parameters of Section 4.1.1? Some training data are required and would conveniently be as subimages containing single objects. One method is to digitize the outlines of the objects by hand but if many objects are available this is laborious. A simple automated method may be worth employing. The prior can then be updated later by successful fits from the more sophisticated Bayesian procedure.

For example, in the study of mouse vertebral shape (Johnson *et al.*, 1985), approximately 500 images of each of the first and second thoracic vertebrae were available. Approximate outlines were extracted by a simple thresholding method, to give a series of about 300 points per outline. After some smoothing these points could then be used to estimate the prior of Section 4.1.1. How are the parameters estimated? Johnson *et al.* (1985) used a Fourier series approach but alternative methods were later sought as local shape differences were difficult to describe.

Alternative approaches for specifying prior distributions include using labelled landmarks on each object, i.e. there is some correspondence, either geometrical or biological, between landmark i on one object and landmark i on another. Examples include the principal components models of Cootes *et al.* (1992) and Kent (1994), smoothed principal components used by Mardia *et al.* (1994) or the use of simple covariances structures in offset normal models in size-and-shape or shape space (Dryden and Mardia, 1991, 1992). If using the offset normal models, inference can proceed by using maximum likelihood.

Again the prior model can be estimated from training data. Rather than hand digitization we used a semi-automatic method for the mouse vertebrae (Mardia, 1989). After obtaining the smoothed outline the landmarks were chosen at important curvature extrema by applying an iterative splitting algorithm for polyline fitting (e.g. Duda and Hart (1973)) to the subset of curvature extrema. Equally spaced pseudolandmarks can then be placed along the outline in between the landmarks. The algorithm is only semi-automatic as a human observer is required to check that the smoothing and the final fit are satisfactory.

The advantage of the labelled natural landmark-based models for the mouse vertebrae example is in the interpretation of local shape differences between different groups. Labelled landmark methods are not always appropriate; for example it does not seem possible to obtain corresponding landmarks on mitochondria. Also, equal numbers of points are required on each figure for the labelled landmark method. In some of the figures the mitochondria appear to be fitted by different numbers of points. Does the method of the paper require equal numbers of points on each figure?

B. D. Ripley (University of Oxford): This meeting occurs during a six-month period I am spending at the 'Computer vision' programme at the Isaac Newton Institute in Cambridge, and it has been our privilege to have both Professor Ulf Grenander and Professor Mike Miller for (separate) short visits. Many of us have been inspired by Ulf's work over many years in seeking realistic models of the contents of an image, of which this paper is an exemplar. It gives me the opportunity to pay tribute to Ulf's seminal contributions—they are often hard to read but reward greatly persistent study by revealing his remarkable intuition.

The section on 'related work' is rather cursory, and I do want to draw the attention of statisticians to some of the excellent work being done on statistical inference for images and video sequences within the computer vision community, very close to the spirit of this paper. In particular, I want to commend to you the volume edited by Blake and Yuille (1992). At the meeting I showed examples on locating cell membranes in electron micrographs, and on tracking of facial features, a buggy and an F-18 Hornet aircraft, all taken from that book. The computer vision community continue to build on their successes and are seeking the collaboration of statisticians, for which we are grateful and look forward to fruitful interactions.

David Mumford (Harvard University, Cambridge): This paper introduces an intriguing algorithm which combines the small stochastic steps of simulated annealing algorithms with large 'jumps' to produce a Markov chain sampling a complex Gibbs field of the type encountered in vision problems. The paper is remarkable in coming to grips with the necessity of making large jumps which change the topology of the sought-for pattern, while also making small incremental improvements with a fixed topology. Shah and I proposed something of this kind (Mumford and Shah, 1985) but had been defeated by its complexity.

One noteworthy antecedent is the work of Brandt *et al.* (1986) on a multigrid algorithm for sampling the Ising model with external field. In fact, their random variables, the spins $x_\alpha = \pm 1$, are equivalent

to those of the authors when there is only one type of organelle and only cyclic graphs. Brandt *et al.* (1986) employed a hierarchy of moves, in which increasingly larger square blocks of pixels are flipped. In the authors' situation, this could introduce a new organelle or radically reshape the partition, chopping a big block out of an organelle, etc. Brandt's method, however, *does not accept or reject a move immediately* but works back down the hierarchy, improving the result with smaller flips, before deciding.

This raises an issue in the present paper which is not stressed: it seems to be computationally intractable to calculate the transition probabilities $Q(x, \mathrm{d}y)$ because of the need to sum over *all* possible new curves $y(1)$ that might be introduced. These new curves can only be sampled, and the illustrations suggest that, even more drastically, only small circles were considered. It is exactly the need to sample well the more probable new curves that forced Brandt *et al.* (1986) to a complex algorithm which delayed acceptance or rejection.

A logical route is to mimic *genetic algorithms* and, rather than to entertain one or a series of global moves on *one* sample, to consider a small population of samples simultaneously. The moves are now the stochastic evolution of the individual members of the population, and the splicing of parts of one sample with parts of another. In the authors' paper, this is especially simple: we may combine two patterns $x(m_1)$ and $x(m_2)$ by taking k_1 of the objects $c(j)$ in $x(m_1)$ together with k_2 of the objects in $x(m_2)$ and forming a new pattern $x(k_1 + k_2)$ out of their union. I believe that this will often be faster and more effective.

Maria Petrou (University of Surrey, Guildford): I would like only to make one comment concerning the four tasks of inference in pattern theory. To me, the fourth task, that of creating regular structures to represent knowledge, is *the* task that bears relevance to all the other three tasks mentioned by the authors. It is this that has defeated us so far, even for the least challenging tasks of all: image restoration. Let us consider, for example, a random texture pattern that has to be restored. Let us assume that only one texture is present in the image, so that the need to introduce model discontinuities and object boundaries is avoided. Random textures can be 'successfully' modelled by Markov random fields (MRFs). The posterior distribution can be derived and the image can be restored, say by using the method of simulated annealing. The restoration can also be done by a multiresolution approach, which when performed carefully can preserve all the implicit and explicit correlations of the model. The idea is to create a pyramid of images of decreasing dimensions and each time to choose the Hamiltonian function describing the image by computing the renormalization group transformation (RGT). Let us also assume that the RGT equations can be solved exactly each time. If I reduce the size of the image to 2×2 and optimize the Hamiltonian of these four pixels and propagate the solution to the finer levels of resolution, will I obtain the correctly restored image? I believe that the answer is *no*! The reason would be that, even though I took pains to record carefully the implicit and explicit correlations of the MRF model, these correlations were incorrect because the model was not sufficiently good. And it would not have been sufficiently good even if the texture had been created by simulating an MRF with chosen parameters. Indeed, in such MRF simulation algorithms, we stop the process when the cost function stops changing noticeably. If, however, we allow the algorithm to carry on, we may end up with a very different looking texture, even though the change of the cost function was extremely slow (Picard, 1991). Fig. 9 shows two simulated images with identical values of their Hamiltonians (from Petrou (1993)). So, what we understand by the word 'model' does not capture all aspects of the object modelled. That is why fractal algorithms have been extremely successful in simulating natural scenes in computer graphics, but they have failed miserably in modelling textures. Our failure lies in the fact that we usually model the result, i.e. the object that we are interested in, whereas we possibly should model the process by which this result was reached.

Lisa Wiffen (University of Leeds): In their paper, Grenander and Miller discuss pattern recognition as a method of automatically locating mitochondria on an electron micrograph. I would like to talk about automatic identification techniques in relation to another particular application: the automation of chromosome analysis.

The data consist of digitized images of cells viewed under an optical microscope. The cells have been exposed to mutagenic chemicals, which can cause abnormalities in the chromosome configurations. The aim is to locate all the chromosomes in an image automatically and further to classify each chromosome as normal or abnormal. If the chromosome is abnormal, it can then be categorized as a specific abnormality.

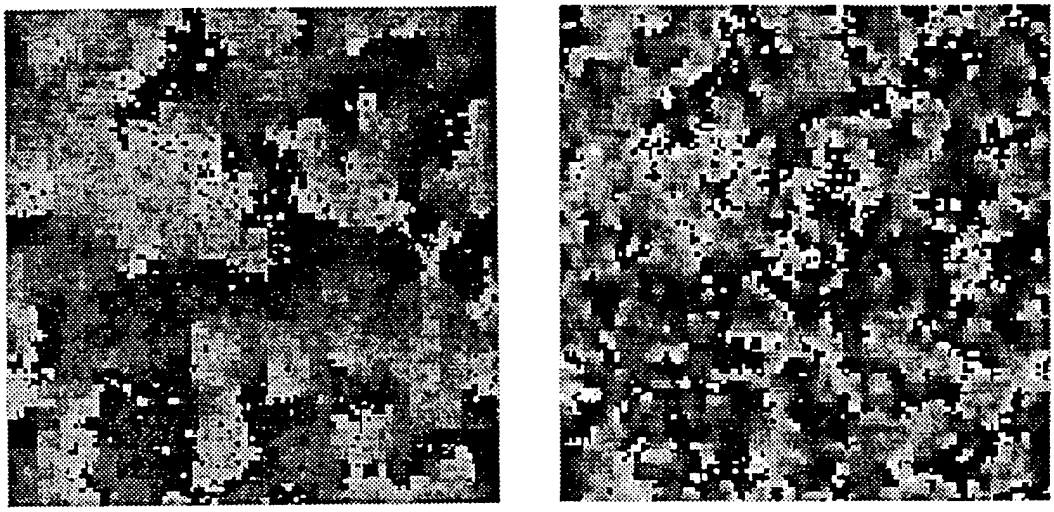


Fig. 9. Two differently looking MRFs described by the same Markov parameters and having Hamiltonians with the same value

The location of chromosomes in an image can be compared with the problem of locating mitochondria, studied by Grenander and Miller. In fact, it may be possible to apply their algorithm to the location of chromosomes. However, such a sophisticated technique is not required. As there is a reasonable contrast between the background of the image and the chromosomes, simple thresholding is sufficient to segment the images. Further, knowledge of the outline of a chromosome does not itself provide enough information about the image to classify it as normal or abnormal. Instead, it is necessary to examine the chromosome image in greater detail.

Normal chromosomes have a characteristic X-shape. The crossover of the X is a constriction in the chromosome arms and can be located anywhere along the length of the chromosome, even at the end. Individual chromosomes are characterized by the length of their arms and the position of this constriction along them. Abnormal chromosomes do not conform to these characteristics, but they do share some features with normal chromosomes; see Fig. 10. Cytologists recognize normal chromosomes by looking

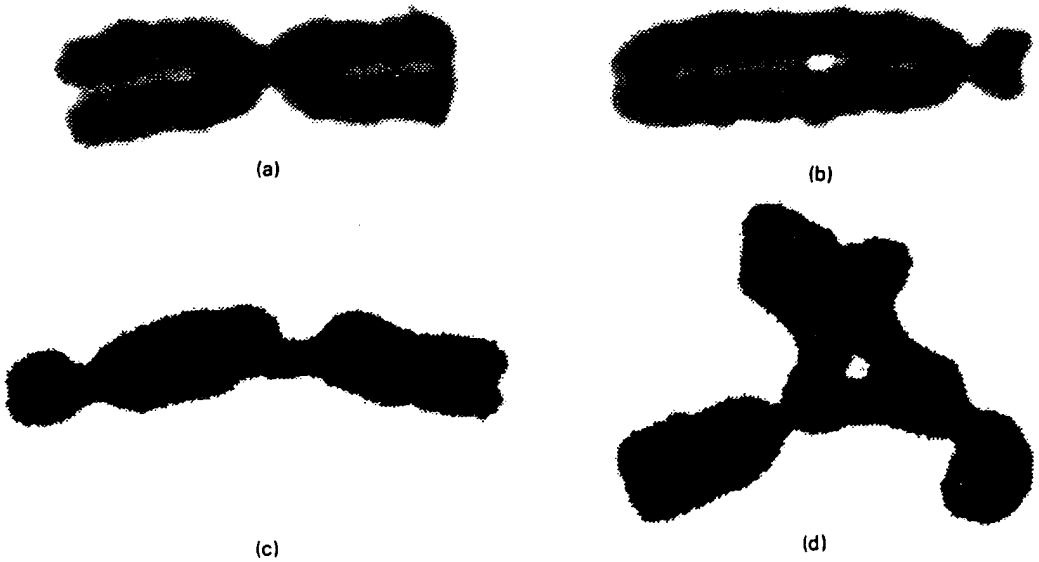
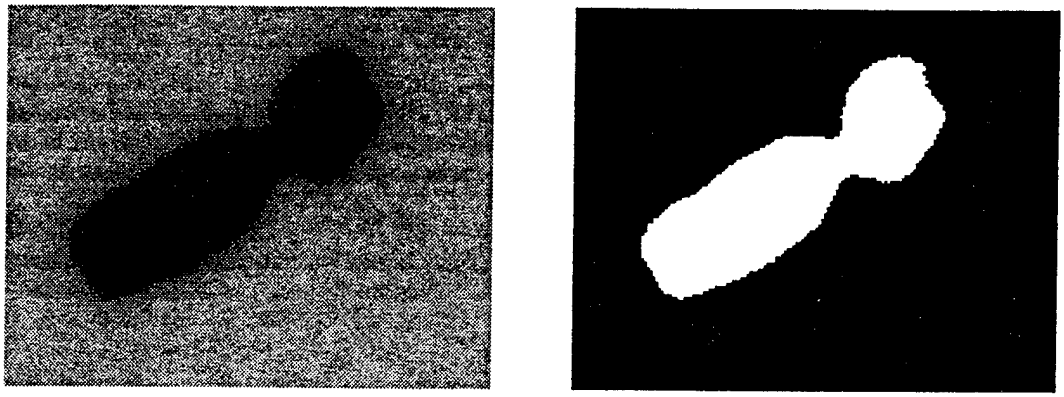


Fig. 10. (a), (b) Normal chromosomes with the constriction in different places; (c), (d) abnormal chromosomes



(a) (b)

Fig. 11. (a) Grey level image and (b) binary image of a normal chromosome

for the two arms and the constriction along them. It is not usually possible to use this approach directly, as thresholding often does not separate the arms of chromosomes; see Fig. 11. Instead, image understanding is implemented in a different way.

The medial axis of a normal chromosome can be found from the minimum of bimodal and the maximum of unimodal grey level profiles across the chromosome. A medial axis found by this method is smooth and continuous for a normal chromosome. There should also be a single cluster of consecutive unimodal profiles, corresponding to the constriction in the chromosome arms. So far, this approach has been very successful in practice.

P. J. Green (University of Bristol): I shall comment here only on the jump–diffusion dynamics, providing the computational engine for the fascinating vision applications in the paper.

Gibbs-sampler-like dynamics like those of corollary 1 will rarely be useful; Markov chain simulation is unnecessary if jumping into a subspace in equilibrium is possible. Much more generally, jumps must be controlled by Metropolis–Hastings acceptance–rejection decisions, and here I shall describe a framework for doing this, discussing only the reversible case. Sufficient conditions are given in theorem 1, but the authors' specific instances of jump measures provide little guidance for the reader with a new problem.

Working in discrete time, consider a family of possible 'moves', indexed by m, counting, for example, a birth from k to $k+1$ objects in the mitochondria example the same move as a death from $k+1$ to k. When the current state is $x \in \mathscr{G}$, the probability of proposing a move of type m and landing in dy is $r_m(x, \mathrm{d}y)$.

Some dimension matching is necessary. Writing ν_m for the measure on $\mathscr{G} \times \mathscr{G}$ given by

$$\nu_m(A \times B) = \int_{x \in A} \int_{y \in B} \mu(\mathrm{d}x)\, r_m(x, \mathrm{d}y),$$

suppose that, for each m, $\nu_m(A \times B)$ and $\nu_m(B \times A)$ have finite positive densities with respect to the same *symmetric* measure ξ_m on $\mathscr{G} \times \mathscr{G}$.

Then

$$\alpha_m(x, y) = \min\left\{1, \frac{\mu(\mathrm{d}y)\, r_m(y, \mathrm{d}x)}{\mu(\mathrm{d}x)\, r_m(x, \mathrm{d}y)}\right\}$$

is well defined (the ratio is that of the two aforementioned densities). This is similar to the usual Hastings ratio, though here the r_m are improper distributions, and on different subspaces. The proposed move is accepted with probability $\alpha_m(x, y)$; otherwise x is retained. The resulting jump measure is

$$q(x, \mathrm{d}y) = \sum_m \alpha_m(x, y)\, r_m(x, \mathrm{d}y)$$

Moves within subspaces can be expressed in the same notation.

It is clear that the resulting chain satisfies the detailed balance condition

$$\int_A \mu(\mathrm{d}x) \int_B P(x \to \mathrm{d}y) = \int_B \mu(\mathrm{d}y) \int_A P(y \to \mathrm{d}x).$$

This framework covers fixed parameter space Markov chain Monte Carlo methods, as used in general Bayesian inference and pixel-based imaging, the birth–death simulations used in point processes and the examples in the paper.

For a simple concrete example, suppose that the target distribution has just two components, with probabilities p_1 and p_2, a univariate density $\pi_1(x)$ on $[0, 1]$ and a bivariate density $\pi_2(x_1, x_2)$ on $[0, 1] \times [0, 1]$. We might consider proposals summarized by

Move type	$x \to$	$(x_1, x_2) \to$
1	(x, u)	x_1
2	(u, x)	x_2
3	(u_1, u_2)	u
4	$(x-u, x+u)$	$\frac{1}{2}(x_1+x_2)$

where the us are independently uniformly distributed on the appropriate intervals. If the probabilities of each move type are bounded away from 0, the dimension matching requirement is met, although the proposal densities have dimensions that are not determined solely by those of the take-off and landing subspaces. The $\alpha_m(x, y)$ are easily derived.

Finally, note that one could contemplate the invention of additional subspaces and associated models, simply to facilitate mixing.

Sophia Rabe (University of Leeds): I would like to focus my contribution on the shape variability present in a laser scan of a human head. This is a more structured problem than the mitochondria considered by Grenander and Miller. The laser scan is an array of measured three-dimensional co-ordinates of points on the surface of the head. One way of representing the shape of the head is to segment the surface into patches of different surface types, based on the local principal curvatures, with peaks corresponding to negative principal curvatures, saddles to principal curvatures of opposite signs, etc. Colour-coded segmentation helps to visualize the shape and is a suitable representation for comparing the faces of patients before and after plastic surgery (Coombes *et al.*, 1991). This analysis suggests that local curvature is an important facial attribute.

Another way to summarize this type of data is to fit a smooth surface with a few parameters. For simplicity, consider the problem in two dimensions using the face profile in Fig. 12 taken from a vertical

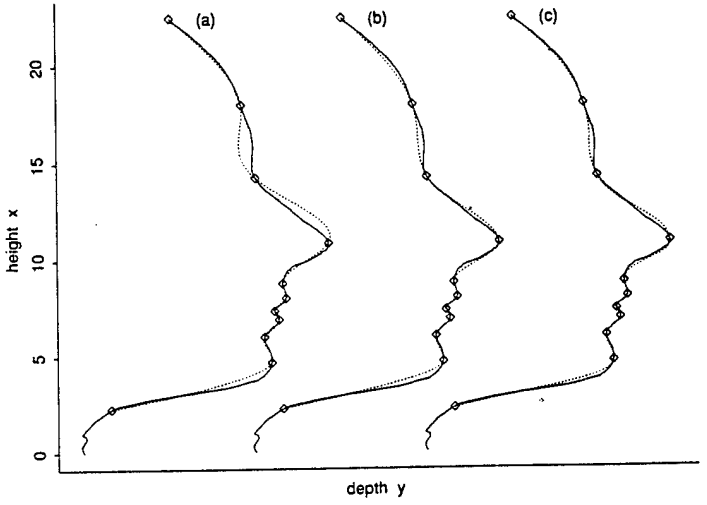


Fig. 12. Face profile (———) with sites (◇) and the following splines (· · · · · ·) superimposed: (a) $r=1$, $2\alpha=3$, $q=0$; (b) $r=1$, $2\alpha=3$, $q=1$; (c) $r=2$, $2\alpha=5$, $q=2$

slice through the laser scan of a female head. Let the height and depth be the explanatory variable x and the response variable y respectively. From the work of Mardia *et al.* (1993), presented in the contribution by Professor Mardia, we can incorporate local curvature and gradient information as follows. Define a spline or kriging predictor $\Phi(x)$ which matches not only the profile depths y_i at n sites x_i, $i=1, \ldots, n$, but also the first q derivatives y_i^ν say, $\nu=1, \ldots, q$,

$$\Phi(x)=\sum_{j=0}^{r} a_j x^j+\sum_{i=1}^{n}\sum_{\nu=0}^{q}(-1)^\nu w_i^\nu \sigma_\alpha^\nu(x-x_i), \tag{39}$$

where $\sigma_\alpha^\nu(x)=\mathrm{d}^\nu|x|^{2\alpha}/\mathrm{d}x^\nu$ for non-integer $\alpha>0$, $r\geqslant[\alpha]$ and $q<\alpha$. The case $r=1$, $\alpha=\frac{1}{2}$, $q=0$ yields the usual interpolating cubic spline.

This cubic spline was fitted to the data at the sites marked by diamonds in Fig. 12, giving the dotted curve (a). Graphs (b) and (c) in Fig. 12 show the same profiles as in (a) with superimposed splines, the dotted curves matching the sites in the gradients ($q=r=1$, $\alpha=\frac{1}{2}$) and curvatures ($q=r=2$, $\alpha=\frac{1}{2}$) of the depth respectively. α must be increased from $\frac{1}{2}$ to $\frac{1}{2}$ to accommodate second derivatives. There is a steady improvement in the fits from (a) to (c) as q and the number of parameters for the spline increase. It may be possible to achieve a better balance between compactness and precision by including different numbers of derivatives at each landmark, using q_i instead of q in equation (39). A three-dimensional model of the face may be constructed by using a generalization of equation (39) to two-dimensional variables x. This is joint work with Professor Kent and Professor Mardia.

A. C. Atkinson (London School of Economics and Political Science): In the talk we were shown a video of pink simulated mitochondria and the progress of the algorithm in identifying them. This made more explicit some of the properties which might be inferred from the still pictures of Figs 3–6.

Templates are born circular and are then gently deformed until they take up the shape of the object that is being identified. It was very noticeable, when viewing the process in time, that objects in the centre were identified first, those cut off by the edges of the frame being identified much later, if at all. The problem presumably is that the intersection of a mitochondrion with the boundary causes a sharp corner, with such violent deformation resisted by the template.

It is not clear to me how the mathematics of the algorithm allow for the fact that some boundaries are edges and may require abnormal deformations. With, for example, eight whole and four cut mitochondria, severe edge effects seem likely.

Chromosomes were mentioned in the discussion. Those shown looked like two short pieces of basically parallel string joined by a knot or two. They would require templates with sharp corners. The edge of the frame should not then introduce any new feature requiring abnormal deformation. It would seem likely that any edge effects will be much reduced when identifying chromosomes, provided that a suitable deformable template can be found.

Julian Besag (University of Washington, Seattle): It is a particular pleasure to add my congratulations to Professor Grenander and Professor Miller. It is already clear that the concept and the implementation of *pattern theory* (e.g. Grenander (1983)) have provided tremendous pay-offs, not only in image analysis but also in general Bayesian computation, where the Markov chain Monte Carlo method has had such a liberating effect.

A feature of the paper, and of some others listed in the references, is the use of a diffusion process to drive the inference machine. In this case, it is a jump process, whereas previous papers have employed basic Langevin diffusions. The simpler versions may also prove to be useful in non-spatial statistical settings, where they can be tightened up as below.

Let $p(s)=\pi(s|y)$, $s\in\mathscr{R}^n$, denote a posterior density for parameters s, given data y. Then the Langevin equation (14) has stationary distribution p and suggests a discrete time Markov chain Monte Carlo algorithm in which the current state s is replaced by

$$s'\sim\mathscr{N}\{s+\tau\nabla\log p(s),\ 2\tau I_n\},$$

where τ is some small positive constant. In this form, which goes back at least to Parisi (1981) in the physics literature and to Grenander (1983) in pattern theory, p is only approximately maintained. However, if instead one uses s' merely as a Hastings proposal for the next state, then the usual acceptance probability ensures that p is an *exact* stationary distribution of the modified Markov chain. Note that

τ is arbitrary and should now be chosen to ensure appreciable proposal increments, accepted moderately often, rather than sufficiently small to mimic closely the Langevin diffusion itself. One might apply such an algorithm to fixed or random subsets of the parameters and τ might be treated as an auxiliary variable with a distribution of its own. Non-Gaussian proposals are also permissible. Sometimes the algorithm is not directly applicable (e.g. for non-negative s) but can be resurrected by a simple (e.g. logarithmic) transformation of the variables. It is not yet clear whether the above Langevin–Hastings algorithm has advantages over standard Hastings methods, such as the Metropolis algorithm and the Gibbs sampler, the first of which can also often be used with vector proposals. Unfortunately, implementation of the Langevin–Hastings algorithm requires knowledge of p up to scale and not just $\nabla \log p$, which presumably precludes it from the applications in the paper.

The following contributions were received in writing after the meeting.

Yali Amit (University of Chicago): The ideas presented by Grenander and Miller offer an excellent conceptual framework for the description of complex systems. I would like to use this opportunity to speculate on some new directions which should be explored in this context. Using Grenander and Miller's terminology, the nodes of the object graphs that they employ are edge elements which are local features in the image. The graph ensures local connectedness and closedness of the boundary. We could, however, introduce a much richer family of local features, in particular various local topographies, of which the edges are only one special case. We would then provide likelihood models of the data in the neighbourhood of the local features. These local features could be arranged in graphs to describe intermediate features such as a boundary segment of high curvature, or a point of overlap of two objects, t-junctions of edges, etc. The intermediate features could be arranged in a graph describing the objects and their possible relationships to each other in the plane.

The bonds in the graphs could enforce connectedness, but they could also enforce a geometric or topological relationship between two nodes or more, e.g. relative location of nodes, relative distances, angles between various directions associated with the local features. At all levels prior probabilities would be assigned to the various values that the graph bonds can assume.

A preliminary attempt at more structured graphs in the context of grey level images can be found in Amit and Kong (1993), where X-rays of hands are analysed. Here the generators are certain local maxima of the image, corresponding to high bone concentrations near the joints and at the palm. The graph describes the *planar and geometric configuration* of the generators (Fig. 13(a)). A penalty penalizing deviations of shape is imposed on the cliques of the graph which are all triangles. Other hand X-rays are analysed by scanning for local maxima (Fig. 13(b)), and then finding the match of the template graph to a subset of these local maxima, such that the sum of the penalties on the individual cliques is minimized (Fig. 13(c)).

Fred L. Bookstein (University of Michigan, Ann Arbor): In Section 6.1 Grenander and Miller diffuse one picture **T**, the *text-book*, on top of another, the *study* **S**, to minimize a combination of elastic energy and squared pixel-by-pixel difference after registration. The warping that results has no convenient analytical form and no 'features'—no straightforward expression in terms of a small number of parameters. Further, the example shown in Fig. 7(f) seems very unfortunately kinked all around the edge of the skull.

If we augment the available data by *landmark points* characterizable both pictorially and biologically, then these and other infelicities of the Langevin diffusion approach can be circumvented by methods that have already appeared in neuroanatomical and statistical publications. The relationship of the landmarks of **S** to those of **T** is borne in a finite dimensional parameter space, David Kendall's *shape space* (Kendall, 1984; Bookstein, 1986; Goodall, 1991). In biometrical applications there is a privileged basis for this shape space closely according with the original task of fitting deformable templates. Let the authors' elastic energy $E(\vec{u})$, equation (37), be replaced by a similar formula related to the bending energy of an infinite, uniform metal plate (Timoshenko and Woinowsky-Krieger, 1959):

$$E(\vec{u}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \int_{\mathbf{R}^n} \left(\frac{\partial^2 u_i}{\partial x_j \, \partial x_k} \right)^2 .$$

The interpolant $\vec{u}$ minimizing E while mapping **T**'s landmarks to those of **S** is linear in the co-ordinates of the landmarks of **S**, and the value of E at the minimum is a quadratic form in those same landmark

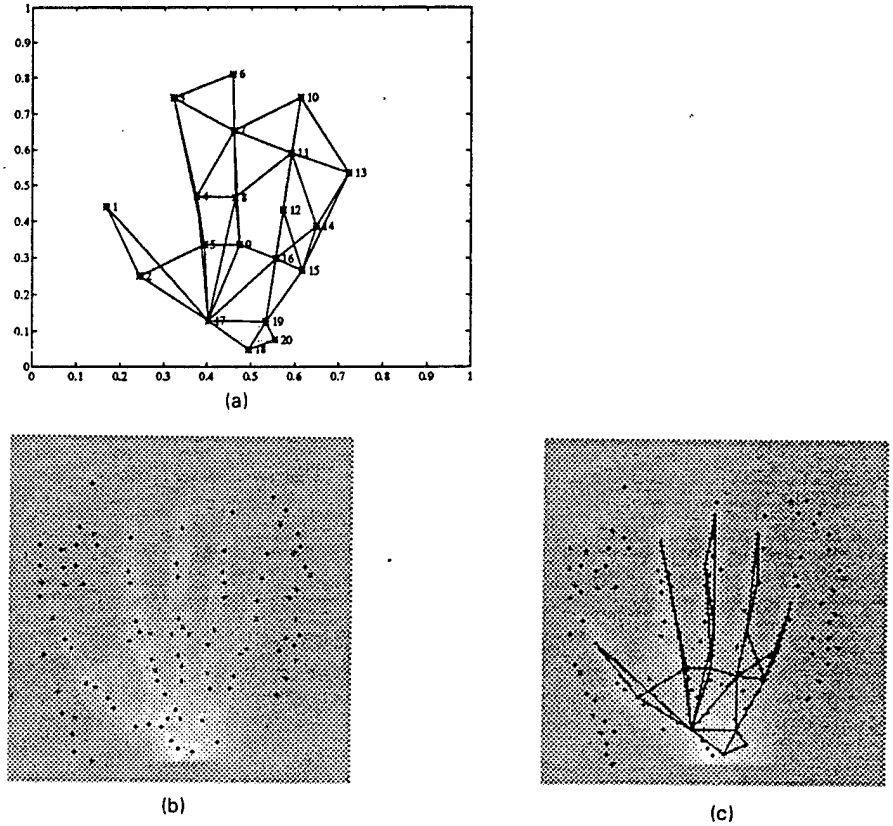


Fig. 13. (a) Template graph, (b) extracted local maxima and (c) optimal match

co-ordinates. Each eigenvector of this form then specifies a 'partial warp' having well-defined localization and geometric scale. In this way the fitted deformation decomposes automatically into a finite set of features potentially of use for both multivariate statistical analysis and display. For $n = 2$ (plane data), Bookstein (1991) demonstrates this formalism, the method of *thin plate splines*, in detail. W. D. K. Green and I have extended the method to incorporate explicit information about boundaries as well (Bookstein and Green, 1993).

The problem set-up in Section 6.1 is incomplete as a 'representation of knowledge' in that the biologist's knowledge of the scientific importance and/or reliability of various features has not been coded. In practice, for certain points and arcs the correspondence between **T** and **S** is far more stable or salient than for others. Text-book figures label and interpret these loci, and specialized algorithms for their automatic extraction are urgently under development in many medical imaging centres. Together, landmarks and edges support a fully developed multivariate praxis (group means, covariances of form, diagnosis, etc.) at no cost to the quality of image matching achieved.

Colin Goodall (Pennsylvania State University, University Park): I enjoyed reading this paper. In the exposition, the authors have been unusually careful in relating to references on Markov random fields.

I wish to articulate some of the similarities and differences between shape theory and the deformable template–jump diffusion approach. The point of view follows from Goodall (1991), which develops the statistical analysis of shape, building on Procrustes analysis (Gower, 1975) as well as the cited work of D. G. Kendall and F. L. Bookstein.

Shape theory addresses the analysis of structures in images

Like Grenander's pattern theory, shape theory is cognisant that important problems of statistical analysis follow image restoration and pattern recognition: we are in the domain of image understanding,

both internal and external, but, for shape analysis, a typical application involves building a statistical model for one or more samples of shapes.

Shape theory involves quotients by a group of similarities

Shape theory considers the analysis of geometrical figures modulo a transformation group, which may be Euclidean, special similarity, affine or projective. The figures may be in ordinary Euclidean space, but need not be (Kendall, 1989, 1992). The group $\mathcal{S}^n$ is more flexible still and allows an exact superposition of two figures, albeit with some penalty, $-\log p(c)$, that may itself be a function $p(c) \propto \exp\{-d(c, c_0)\}$ of the distance d between shapes.

Morphometrics builds on existing scientific expertise

Anatomists can identify homologous features—landmarks—across images with high accuracy. In biology, the analysis of shape follows a notably interdisciplinary approach.

Through the analysis of landmark data, the statistical analysis of shape becomes an extension of multivariate analysis. Some theoretical aspects are explored by Goodall and Mardia (1993). Most important, the same basic principles of statistics apply to the analysis of shape as in an introductory statistics course. Shape is to be thought of as a dependent variable in statistical models, as outlined in Goodall (1993), and it simply remains to explain how to accommodate transformation groups that are more general than the translations of randomized block designs.

The statistical analysis of shape, therefore, represents a classical approach to image understanding, in contrast with the computational Bayes approach of Grenander and Miller. Shape models will become increasingly complex, incorporating edge and textual data as well as landmarks, and including a variable number of objects in a scene. Then the distinction blurs, as the computational Bayes approach becomes the most viable strategy. Also, until now, model identifiability has been explicit; images in $\mathcal{I}$ can be identified directly with configurations in $\mathcal{L}(\mathcal{R})$, and the similarities of $\mathcal{S}$ are accommodated in explicit nuisance parameters.

I look forward to continued development in both shape theory and Grenander's pattern theory, to mutual challenges and to further cross-fertilization, e.g. as in Mardia *et al.*'s (1991) use of shape distributions as shape priors in deformable template models.

Jim Kay (University of Stirling): I wish to congratulate the authors on their impressive achievements as described in this paper. It is most interesting to see these further implementations of Professor Grenander's seminal work in image analysis and pattern theory. I believe that the methodology of the paper is relevant to the following problem, although there are potential complications.

Parasitologists wish to develop an automated image analysis procedure for the identification of the parasite *Gyrodactylus salaris* and its discrimination from other species of *Gyrodactylus*. This pathogen is known to have been responsible for the eradication of populations of wild salmon in Norwegian rivers and it is desired that monitoring stations in the UK can detect its presence on salmon. A largish database of images of the various species of *Gyrodactylus* is available, having been obtained by using both scanning electron microscopy (SEM) and light microscopy (LM). The differences between the species are subtle, except to one of the few world experts, and involve the shapes of the hooks that are used by the parasites to attach to the host salmon. For each specimen one could hopefully, and automatically, extract the shape of its hooks and exploit the fact that there is greater between-species variability of shape than within-species variation to classify a specimen. However there are two complications. Firstly, owing to cost and availability, it is required that the system uses LM images which have a coarser resolution than that of SEM images. However, the availability of a training set of images in both imaging modalities means that some kind of shape calibration might be performed. Secondly, the shapes of the hooks depend on covariates such as the temperature of the water (time of year) and its salinity. In particular, seasonal effects can be quite dramatic. Note that the data are collected cross-sectionally rather than longitudinally and so approaches such as tracking would not seem to be directly applicable. Thus it would seem that the classification procedure will require to be performed conditionally on the values of these and possibly other covariates. Would the authors be prepared to offer any suggestions?

Andrew Lawson (Dundee Institute of Technology): I would like to make the following comments on issues relating to object recognition and the use of prior information. My comments echo those of Professor Baddeley and Marie-Colette van Lieshout in that prior distributions from stochastic geometry are of great use in object modelling. Both Poisson process and Markov process object models can describe a rich class of spatial structures with relatively simple parameterization. For example, Markov line process

models can be used to model linear features found in the mitochondria examples. In some situations the clustering tendency of objects is of prime importance: such features as cluster centre estimation and membership labelling can also be modelled under Poisson cluster or Cox process models. Markov chain Monte Carlo methods for object models have been discussed by Baddeley and van Lieshout (1992b, 1993) and, specifically, for the case of clustering by Lawson (1993) and Lawson *et al.* (1993). For example, cluster centre estimation for the Neyman–Scott cluster model can be based on the posterior likelihood ratio

$$\frac{p(\mathbf{x}\cup\{u\}|\mathbf{y})}{p(\mathbf{x}|\mathbf{y})}=\rho\prod_{j=1}^{m}\left\{1+\frac{h(y_i-u)}{\sum_{i=1}^{n}h(y_j-x_i)}\right\}\exp\left\{-\int_U h(v-u)\,dv\right\},$$

where $h(\,)$ is a radial cluster distribution function, ρ is the parent rate and u is a proposal centre, $\mathbf{x}$ is the current configuration of n centres and $\mathbf{y}$ are the m data points.

This can be included in a Metropolis–Hastings step.

David Phillips (Imperial College of Science, Technology and Medicine, London): The authors are to be congratulated for presenting an illuminating exposition of state of the art modelling and inferential algorithms for use in image analysis.

As noted in the paper, realistic image scenes may be composed from the superposition of many shapes (see the mitochondria images) which can be characterized by using suitable deformable templates. An

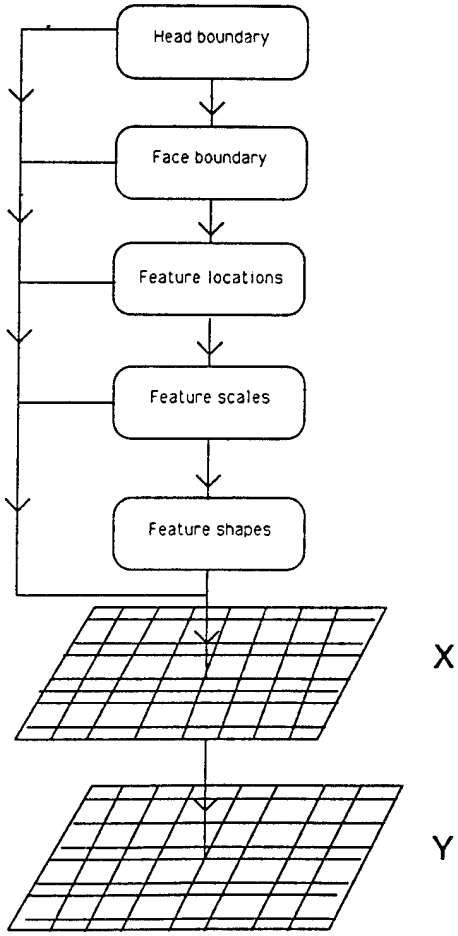


Fig. 14. Graphical model

appealing way to model the arrangement and interaction of these shapes, when they are known, is to embed the templates in a hierarchy, exploiting the interplay between local and global features present in many types of image.

An example described in Phillips and Smith (1993) relates to models of face images, for which the hierarchical structure can be schematically represented in the form of a graphical model (see, for example, Spiegelhalter *et al.* (1993)), as illustrated in Fig. 14, where **X** denotes the segmented image and **Y** the observed image.

The use of hierarchical models based on a number of related modelling stages, and exploiting conditional independence between levels, is an attractive way of approaching many imaging problems requiring a combination of local and global prior specification, which has the advantage that further levels of refinement, if required, may easily be added to the model.

Michael W. Vannier (Mallinckrodt Institute of Radiology, St Louis): Characterization of biological shape variation has had little practical importance since methods for rigorous analysis have significantly lagged developments in techniques for the acquisition and manipulation of surface and volumetric data in the last decade. It has not been possible to deal efficiently and reliably with biological shape variations in clinical medicine. Grenander and Miller have given us a new set of tools to understand biomedical silhouettes, projection radiographs, serial slices and volumetric image data sets. An extension to three and four dimensions where shape varies with time is awaited.

An application to populations allows the separation of individual differences from normal variation. Grenander and Miller's work will revolutionize allometry applied to the study of normal and abnormal human growth and development, interspecies variation and sexual dimorphism. Their method provides a mathematically tenable means to isolate markers for testing heritability of traits through quantitative genetic analysis.

Magnetic resonance imaging (MRI) has emerged recently as a major research and clinical modality to study the brain, heart, musculoskeletal system and other organs and body regions. Despite great flexibility in conducting imaging experiments that yield high contrast between body tissues, the specificity for MRI is low. For example, a reliable separation of normal brain tissue components such as grey matter, white matter, cerebrospinal fluid and others is very difficult.

Statistical pattern recognition has been applied with modest success to the segmentation of MR scans. These methods are based on feature extraction from observed measurements categorized according to a rule set. Supervised methods are often required owing to instrument signature and measurement variations so fully automatic segmentation of MR data sets has not been practical. Superior results were demonstrated by Grenander and Miller with an important generalization of statistical pattern recognition to achieve what no other method has been able to do—to increase the specificity of MRI.

Neuromorphometric studies of the brain in subpopulations afflicted with neuropsychiatric disorders are performed *in vivo* by using MRI, based on the premise that symptomatic individuals share regional shape and volume differences which correspond to focal abnormalities. Identification of these sites requires high precision image analysis that cannot be achieved manually, not to mention the tedium of processing large numbers of images. Using the methods of Grenander and Miller, we can automatically scale, register, segment and label complex MR images, given the existence of prior knowledge in the form of an electronic atlas or text-book.

The **authors** replied later, in writing, as follows.

Statistical knowledge representations

In response to Professor Clifford's thought-provoking remarks concerning 'is this really statistics?', 'is it information engineering?', we make an analogy with statistics in the early 20th century. At that time large-scale sample surveys became common with a theory for their design just beginning to appear. What had earlier appeared intractable could now be handled, both because mechanical devices (such as the Hollerith machine) became available and because the theoretical underpinning for the inferences required had been created. Today we are faced with a new situation with powerful new sensor modalities becoming available which make it possible to acquire astronomical amounts of data fast and at rapidly decreasing costs. The user is forced to develop tools for utilizing the data sets. This is not just a matter of developing relevant software, although that is of course needed. We must also learn to represent the knowledge that will scientifically support the inference algorithms, and give a conceptual basis for their formal development. We have argued that pattern theory provides such a framework.

But then, who is going to do it? It is difficult to see how probabilistic ideas can be avoided with the immense variability inherent in biomedical image ensembles. This is familiar territory for the statistician and engineer of the signal processing type as well. The mathematics used in our paper may not be familiar to all statisticians, but the way of thinking about data, whether Bayesian or not, certainly belongs to the statistical field and our hope is that more statisticians will be attracted to this challenging endeavour.

Turning to Professor Besag, an innovator in our field of study, he points out that related methods 'may also be useful in more conventional statistical problems' and we certainly agree. In our paper we have remarked, pointing to several case-studies, that the application to mitochondria micrographs is a special case of a methodology of wide applicability. In all such cases the fundamental difficulty lies in the creation of knowledge representations. This will be true *a fortiori* in the near future when we have to deal with gigabytes of information. The derivation of the inference algorithms from these representations will require considerable effort. Dr Petrou emphasizes that 'the task of creating regular structures to represent knowledge is *the* task'. We agree. It is tempting just to apply generally known statistical principles to the image ensembles without modelling the underlying structure.

Modelling and constructing the prior distribution

Professor Kent's insight concerning the usefulness of the 'mirror symmetry' for the mitochondria prior is fully appreciated. It has not yet been incorporated, although we are currently investigating its applicability and the associated sine–cosine all-real rotation; instead of the cyclic group to characterize invariances we could use the dihedral group. We appreciate his careful examination of the assumptions associated with our complex block diagonalization of the block stationary processes. Concerning simplifying the covariances, the basis representation is used to enforce closure. Flowing through the rotated co-ordinates is not significantly complicated by the non-zero quadratic term that the prior contributes to the posterior.

Dr Dryden inquires into the actual construction of the prior. 497 mitochondria in 41 images were hand traced and sampled for constant arc length, with the Fourier transform means and variances computed. Concerning fixed *versus* variable numbers of arcs, the jump–diffusion reallocates the number of arcs (power of 2 for fast Fourier transform use) to maintain roughly constant arc length.

Jump–diffusion strategies

Professor Mumford and Professor Green provide insight into the jump–diffusion mechanics. Concerning the study of practical choices for the jump sampling two alternative procedures are described. Professor Mumford is quite correct in that the first approach, shown for the mitochondria and membrane examples, involves explicit calculation of the transition probability $Q(x, dy)$ over all possible new boundaries or curve segments. For mitochondria this is crudely approximated by choosing the mean shape from the prior with placement attempted at 64 alternative places biased via the intersection penalty into the uncovered part of the grid. For the membranes new segments are added to existing ones requiring sampling a distribution associated with the single scale–rotation distribution. The second version of jump selection has also been implemented proceeding by drawing from the prior on shape and position and accepting via the likelihood. This can be computed exactly; however, in the electron micrograph application method the first version has been found to be efficient. For the tracking recognition application the second method is used since drawing from the prior on airplane dynamics is fast and effective.

Concerning Professor Green's questioning of the applicability of the Gibbs-sampler-like dynamics, the posterior must be simulated only over the space of configurations in the range of a single jump move. This corresponds to sampling a distribution associated with single objects which can be of low or high dimension. The original posterior involves multiple objects, for which direct sampling is impossible.

Professor Mumford's suggestion to enlarge the space of moves to include stochastic evolution before accepting is intriguing and our colleague K. Mark has explored this in the computational linguistics context of sentence parsing (Mark *et al.*, 1992). For realistic grammars exhaustive chart parsing is a demanding n^3-computation (n being sentence length). Random parsing is performed by using a 'super move' consisting of a series of graph changes pushing randomly towards a complete parse. The moves are not accepted or rejected immediately. The resulting hypothesized complete parse becomes the candidate which is tested on. The connection to genetic algorithms made by Professor Mumford seems intriguing in which the power of sampling multiple choices on any one jump transition could perhaps be obtained while receiving the benefit of diffusion search for spreading candidates through connected parts of parameter space.

Professor Mardia raises an important question concerning systematic search *versus* jump–diffusion. The jump–diffusion algorithm allows us to organize systematically the components associated with a brute force search. When first discussing how the deduction process should proceed we thought that

(a) the algorithm should scan the picture looking for large shapes before focusing in, analogous to *saccades in the visual system*, which correspond to large jumps in visual space, and
(b) the algorithm should be able to break and fuse objects.

The mathematical structure for organizing these transformations are jump–diffusions. The algorithm works equally well irrespective of the number of objects: the global search is obtained from the saccades that it performs via the jump process, the local structure from the diffusion.

Returning to Professor Besag's discussion of the use of a gradient-based proposer, we have been motivated to use Langevin search since the mid-1980s because it is natural for continuous-valued variables and because it supports completely parallel site updating, a desirable feature for parallel machine implementation. Interesting analytic formulae result. For example theorem 3 shows that variation of the posterior with respect to shape parameters corresponds to line integrals around the boundary; in Miller *et al.* (1993) this has been extended to three dimensions with corresponding surface integrals. Dr Petrou comments on our choice of the continuum, and the difficulties which arise with scaling on a discrete lattice. In imaging the lattice is often chosen, although it is often a technological artefact related to the sensor and not the parameter space. The continuum is actually simpler since the Euclidean spaces support natural similarity groups: translation, scaling, rotation, affine groups and projective transformation as well.

Hierarchical graph models

We applaud Professor Amit and Dr Phillips for their latest developments using graph-based templates. The introduction of a hierarchy of graphs with the more general generators forming the patterns is a major emphasis of our work. For example, examine the stochastic phrase structure language models explored in Mark *et al.* (1992). The graphs are trees with loops at the leaves. The generators are the production rules of a probabilistic context-free grammar describing the way in which non-terminals (syntactic variables) may be rewritten. The bond values are the non-terminals and terminals (words). A configuration consists of a set of generators (production rules), placed at the sites of the tree graph type $\Sigma = \text{FOREST}$. To accommodate Markov relationships on the leaves (terminals or words) of the trees, interactions between the words in the lexicon are enforced. Acceptor functions on the bond values are chosen to reflect the stochastic structure of random branching processes (Harris, 1963; Grenander, 1967; Miller and O'Sullivan, 1992) on the context-free tree bases of the derivations and the Markov chain structure on the word (leaves) of the tree.

The parameters for such a model were estimated from a subset of the Penn TreeBank corpus consisting of 1013789 words in 42254 sentences which have been machine parsed and hand corrected using a context-free grammar containing 24111 rules to the preterminal level and 78929 to the word level. Markov leaf structure was estimated from 389440 bigrams and 744162 trigrams in the data. Mark has compared the entropies of four different language models: the mixed graph model, compared with the bigram and trigrams Markov chain models and a purely context-free random branching process model. The power of the mixed graph model is the discriminability that it provides (as measured via the entropy decrease) as a function of the number of parameters. As shown in Fig. 15 the mixed graph model provides dramatic reductions in entropy for a small addition of parameters.

Knowledge representations on the continuum

Professor Goodall, Professor Bookstein and Professor Vannier raise significant points concerning the future of complex knowledge representations. Professor Goodall suggests that 'shape theory will become increasingly complex . . . the distinction blurs as computational Bayes becomes the most viable strategy'. We agree. Indeed, as biologists and engineers develop ever more powerful technologies for image acquisition the analysts will be forced to incorporate huge amounts of information in their representations.

Professor Bookstein suggests that a small number of parameters or *landmarks* are closely linked to *representations of knowledge*. Our work on mitochondria, membranes and anatomy has given rise to representations in the continuum, involving high dimensional vector fields. Specialized landmarks when available (sulci, fissures, etc.) can be accommodated in the methods, and we are currently doing this. As emphasized by Professor Vannier, the advent of new imaging modalities makes it possible to generate massive data sets which accurately represent the 2–3 dimensional geometry in the near continuum. Examine the questions that our collaborators are investigating. Felleman and Van Essen (1991) are

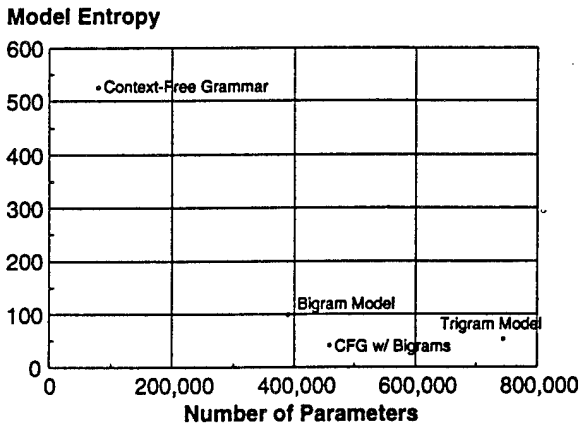


Fig. 15. Comparison of the model entropy of the four language models: the mixed graph model, the bigram and trigram models, and the context-free branching process model

studying the shape and folding of neocortex; Csernansky *et al.* (1991) are quantifying the morphological changes in the brain associated with schizophrenia. Having a precise representation in the continuum seems essential. To illustrate, shown in Fig. 16 are sections from visual cortex taken from David Van Essen's laboratory at an approximate 10× increase in resolution over the magnetic resonance data. Figs 16(a) and 16(b) show slices from two different macaque monkeys; Fig. 16(d) shows a result generated by Christensen *et al.* (1993) of the top left anatomy elastically deformed into the top right. Fig. 16(c) shows the extent of the necessary deformation as applied to the original grid.

Dr Wiffen and Dr Rabe first made us aware of their work at the 1993 Leeds workshop on three-dimensional shape. These are exemplary of what we believe to be an area of growing applications:

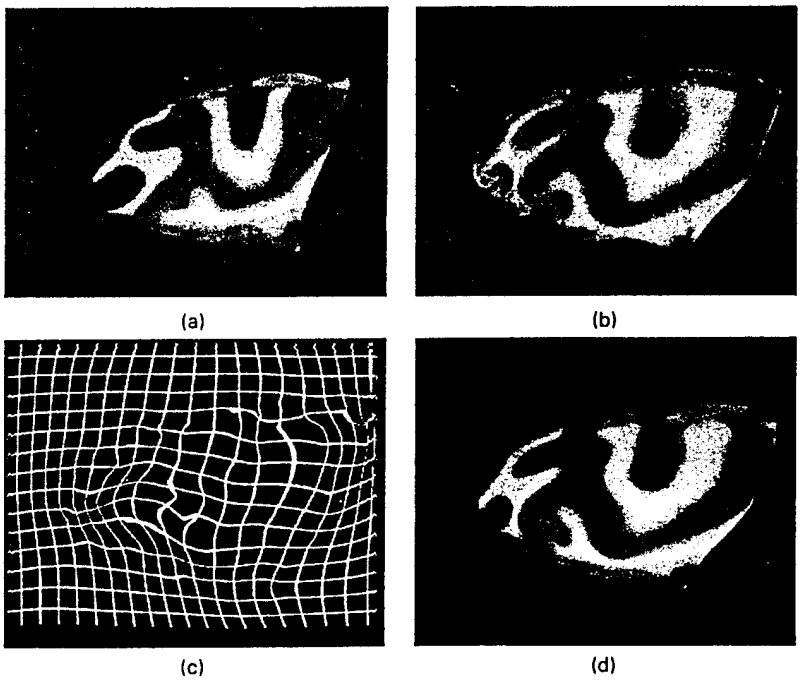
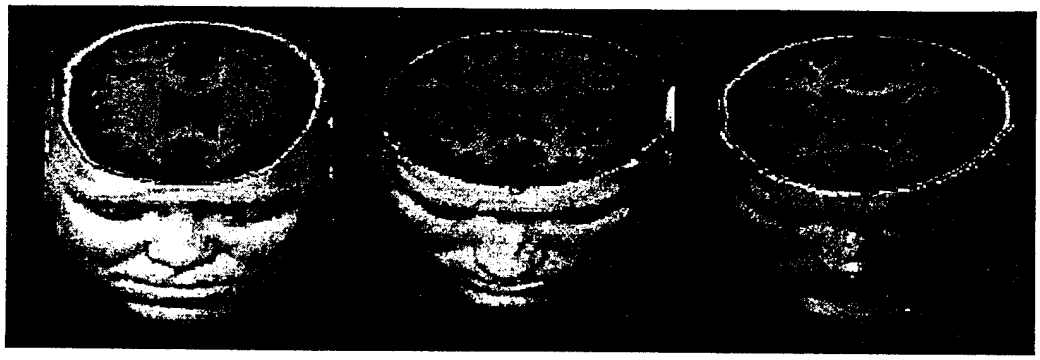


Fig. 16. (a), (b) Two visual cortical slices from two macaque monkeys, (c) the deformed grid and (d) the result of deforming (a) into (b) (data from David Van-Essen and Tom Coogan)



(a) (b) (c)

Fig. 17. Result of elastically deforming (a) a 128^3 voxel, three-dimensional MPRAGE magnetic resonance volume into (b) the individual is shown in (c): notice how similar the magnetic resonance data and the faces are in (b) and (c) (data taken from collaborators Marcus Raichle and Michael Vannier)

the mixture of linear, surface and volume representations. As shown in Fig. 17, we have built our first three-dimensional magnetic resonance imaging (MRI) text-book of the human head including both the soft tissue facial surface, bony skull and brain volume. Figs 17(a) and 17(b) show the facial surface and volumes of two individuals from full three-dimensional MRI volumes generated in Michael Vannier's laboratory. Fig. 17(c) shows a three-dimensional elastic deformation generated by Christensen *et al.* (1994) of the left volume into the middle volume. We are currently examining surface representations such as proposed by Dr Rabe for inclusion in the text-book.

Jump–diffusion strategies

Professor Ripley alludes to the tracking results mentioned in the comprehensive book of Blake and Yuille (Harris, 1992). Automated target tracking and recognition are well-known problems in the vast control and signal processing literature. Multiple-target tracking posed as state estimation is discussed extensively in Bar-Shalom and Fortmann (1988) and Bar-Shalom (1990), including feature-based tracking such as alluded to by Ripley (Harris, 1992). Dynamics-based Kalman filter techniques are emphasized, with *linear equations of state* playing a fundamental role. For situations in which the observed data are non-linear in target parameters the use of the extended Kalman filter has been proposed. More relevant is the growing body of work on tracking from sensor *arrays* based on Schmidt's classical characterization of the so-called *array manifold*. Linear state to data models do not apply. Recognizing this, investigators (Rao *et al.*, 1993; Sastry *et al.*, 1991; Sword *et al.*, 1990) have explored generation of gradient-based estimators of the position (state) at each instant of time from the likelihood; these estimators serve as the measurements in the Kalman filter state equations. Simplifications are again required: targets are assumed stationary with multiple measurements at each sample time (required so that the gradient-based estimates are asymptotically Gaussian (Rao *et al.*, 1993)), and linear models of target dynamics (usually constant velocity–constant acceleration) are adopted. More fundamental is the explicit separation of the tracking and recognition problems.

The approach used here is to couple the recognition and tracking problems reflecting the fact that the rotational and translational motions of rigid bodies as described via the classic set of Newtonian differential equations are coupled. The recognition data provide the orientation information. As shown in Fig. 8 random sampling via the jump–diffusion allows for the conditional mean to be generated from the single unified posterior on the tracking and recognition data. The family of graph changes correspond to increases and decreases in track length and changes in target type. The acceptance–rejection alternative analogous to the second part on theorem 4 was used.

Professor Baddeley, Dr van Lieshout and Dr Lawson point out that various aspects of the random inference algorithm exhibit close resemblance to spatial birth-and-death processes. Closer scrutiny, however, reveals a host of differences and difficulties which we have attempted to solve systematically and precisely. It is suggested that the configurations be viewed as realizations of marked point processes, the marks taking values in Euclidean spaces. Priors, such as the Strauss distribution as used by

Dr van Lieshout, would then become available. For rigid and regular shapes, models which control overlap may be helpful. For *highly deformable biological shape* such as mitochondria it does not appear that a prior distribution based on centres and radii will be informative. Our approach via the intersection penalty is to encode spatial interactions by using a completely connected graph at the object level with areas laboriously painted and calculated for the Gibbs potential. A second point concerns the apparent similarity of our jump–diffusion construction to various allusions to jump processes or diffusions. Examine the reference to Geyer and Møller's Metropolis–Hastings algorithm: *containing a mixture of transition kernels, the first an analogue of the diffusion and the second the jump.* This is misleading as in the aforementioned work the random algorithm simply proposes on fixed times, with no gradient. In the work of van Lieshout (1993) and Baddeley a diffusion component is not included although positions in Euclidean spaces of the rigid objects are parameters of interest. Gradient search through the connected parts of the space seems natural; a mathematical framework (such as our theorem 4) allowing for the state to evolve continuously between random jump times, and to be carried forwards after the jump, seems extremely important. But this also presents technical challenges. The state space is not compact implying that the drifts are not bounded over R^n. Periodizing or reflecting the translation–rotation group does not seem natural to us. So to prove irreducibility and thus uniqueness of the invariant measure we cannot use the standard theorems on the existence of densities for diffusions (as was elegantly done in Geman and Hwang (1987)). In van Lieshout (1993), although interested in R^n, theorems for the jump process are proven in a bounded or discrete subset of R^n. We have been more ambitious. Irreducibility is proven for the diffusion within each Euclidean space by proving that for a new *killed* process (defined by first passage out of compact subsets) a density exists, implying irreducibility within each subspace via the diffusions; irreducibility over the full state space follows from the jumps (see Grenander and Miller (1991) and Amit *et al.* (1993) for technical details).

REFERENCES IN THE DISCUSSION

Amit, Y., Grenander, U. and Miller, M. I. (1993) Ergodic properties of jump–diffusion processes. Submitted to *Ann. Appl. Probab.*

Amit, Y. and Kong, A. (1993) Graphical templates for image matching. *Technical Report 373.* Department of Statistics, University of Chicago, Chicago.

Baddeley, A. J. and van Lieshout, M. N. M. (1991) Recognition of overlapping objects using Markov spatial processes. *Research Report BS-R9109.* Centrum voor Wiskunde en Informatica, Amsterdam.

———(1992a) ICM for object recognition. In *Computational Statistics* (eds Y. Dodge and J. Whittaker), vol. 2, pp. 271–286. Heidelberg: Physica.

———(1992b) Object recognition using Markov spatial processes. In *Proc. 11th IAPR Int. Conf. Pattern Recognition*, pp. B136–B139. Los Alamitos: Institute of Electrical and Electronics Engineers Computer Society Press.

———(1993) Stochastic geometry models in high-level vision. In *J. Appl. Statist.*, **20**, 233–258.

Baddeley, A. and Møller, J. (1989) Nearest-neighbour Markov point processes and random sets. *Int. Statist. Rev.*, **57**, 89–121.

Bar-Shalom, Y. (ed.) (1990) *Multitarget-multisensor Tracking.* Norwood: Artech House.

Bar-Shalom, Y. and Fortmann, T. E. (1988) *Tracking and Data Association.* New York: Academic Press.

Blake, A. and Yuille, A. (eds) (1992) *Active Vision.* Cambridge: Massachusetts Institute of Technology Press.

Bookstein, F. L. (1986) Size and shape spaces for landmark data in two dimensions. *Statist. Sci.*, **1**, 181–242.

———(1989) Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattn Anal. Mach. Intell.*, **11**, 567–585.

———(1991) *Morphometric Tools for Landmark Data.* New York: Cambridge University Press.

Bookstein, F. L. and Green, W. D. K. (1993) A feature space for edgels in images with landmarks. *J. Math. Imgng Vis.*, **3**, 231–261.

Brandt, A. Ron, D. and Amit, D. (1986) Multi-level approaches to discrete-state and stochastic problems. *Lect. Notes Math.*, **1228**, 65–98.

Christensen, G. E., Miller, M. I. and Vannier, M. (1994) A 3d deformable magnetic resonance textbook based on elasticity. *American Association for Artificial Intelligence Symp. Application of Computer Vision in Medical Image Processing, Stanford, Mar. 21st–23rd.*

Christensen, G. E., Rabbitt, R. D. and Miller, M. I. (1993) 3D brain mapping using a deformable neuroanatomy. *Phys. Med. Biol.*, to be published.

Coombes, A. M., Moss, J. P., Linney, A. D., Richards, R. and James, D. R. (1991) A mathematical method for the comparison of three dimensional changes in the facial surface. *Eur. J. Orthdont.*, **13**, 95–110.

Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J. (1992) Training models of shapes from sets of examples. In *Proc. British Machine Vision Conf.*, pp. 9–18. London: Springer.

Csernansky, J. G., Murphy, G. M. and Faustman, W. O. (1991) Lymbic/mesolimbic connections and the pathogenesis of schizophrenia. *J. Biol. Psychol.*, **30**, 383–400.
Dryden, I. L. and Mardia, K. V. (1991) General shape distributions in a plane. *Adv. Appl. Probab.*, **23**, 259–276.
———(1992) Size and shape analysis of landmark data. *Biometrika*, **79**, 57–68.
Duda, R. O. and Hart, P. R. (1973) *Pattern Classification and Scene Analysis*, p. 338. New York: Wiley.
Felleman, D. J. and Van Essen, D. C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebr. Cort.*, **1**, 1047–3211.
Galton, F. (1878) Composite portraits. *J. Anthrop. Inst.*, **8**, 132–142.
Gelfand, S. B. and Mitter, S. K. (1991) Weak convergence of Markov chain sampling methods and annealing algorithms to diffusions. *J. Optimizn Theory Applic.*, **68**, 483–498.
Geman, S. and Hwang, C.-R. (1987) Diffusions for global optimization. *SIAM J. Control Optimizn*, **24**, 1031–1043.
Geyer, C. J. and Møller, J. (1993) Simulation procedures and likelihood inference for spatial point processes. *Technical Report 260*. University of Aarhus, Aarhus.
Goodall, C. R. (1991) Procrustes methods in the statistical analysis of shape (with discussion). *J. R. Statist. Soc.* B, **53**, 285–339.
———(1993) Shape as dependent variable in statistical models. *Institute of Mathematical Statistics Meet.* To be published.
Goodall, C. R. and Mardia, K. V. (1993) Multivariate aspects of shape theory. *Ann. Statist.*, **21**, 848–866.
Gower, J. C. (1975) Generalized Procrustes analysis. *Psychometrika*, **40**, 33–50.
Grenander, U. (1967) Probability measures for context-free languages. Brown University, Providence.
———(1983) Tutorial in pattern theory. Division of Applied Mathematics, Brown University, Providence.
Grenander, U., Chow, Y. and Keenan, D. (1990) *HANDS: a Pattern Theoretic Study of Biological Shapes*. New York: Springer.
Grenander, U. and Miller, M. I. (1991) Jump–diffusion processes for abduction and recognition of biological shapes. *Monograph*. Electronic Signals and Systems Research Laboratory, Washington University, St Louis.
Harris, C. (1992) Tracking with rigid models. In *Active Vision* (eds A. Blake and A. Yuille). Cambridge: Massachusetts Institue of Technology Press.
Harris, T. E. (1963) *The Theory of Branching Processes*. Berlin: Springer.
Illingworth, J. and Kittler, J. (1988) A survey of the Hough transform. *Comput. Vis. Graph. Image Process.*, **44**, 87–116.
Johnson, D. R., O'Higgins, P., McAndrew, T. J., Adams, L. M. and Flinn, R. M. (1985) Measurement of biological shape; a general method applied to mouse vertebrae. *J. Embryol. Expmntl Morph.*, **90**, 363–377.
Kendall, D. G. (1984) Shape manifolds, procrustean metrics and complex projective spaces. *Bull. Lond. Math. Soc.*, **16**, 81–121.
———(1989) A survey of the statistical theory of shape (with discussion). *Statist. Sci.*, **4**, 87–120.
———(1992) Spherical triangles revisited. In *The Art of Statistical Science, a Tribute to G. S. Watson* (ed. K. V. Mardia). Chichester: Wiley.
Kent, J. T. (1994) The complex Bingham distribution and shape analysis. *J. R. Statist. Soc.* B, **56**, 285–299.
Kent, J. T., Mardia, K. V. and Walder, A. N. (1992) Conditional cyclic Markov random fields. *Research Report*. Department of Statistics, University of Leeds, Leeds.
Lawson, A. B. (1993) Discussion on the Royal Statistical Society meeting on the Gibbs sampler and other Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **55**, 61–62.
Lawson, A. B., van Lieshout, M. N. M. and Baddeley, A. J. (1993) Markov chain Monte Carlo methods for spatial cluster processes. Submitted to *J. R. Statist. Soc.* B.
van Lieshout, M. N. M. (1991) A Bayesian approach to object recognition. *Res. Informatics*, **4**, 185–190.
———(1994) Stochastic annealing for nearest-neighbour point processes with application to object recognition. *Adv. Appl. Probab.*, **26**, in the press.
Mardia, K. V. (1989) Discussion on A survey of the statistical theory of shape (by D. G. Kendall). *Statist. Sci.*, **4**, 108–111.
Mardia, K. V., Dryden, I. L., Hurn, M. A., Li, Q., Millner, P. A. and Dickson, R. A. (1994) Familial spinal shape. *J. Appl. Statist.*, to be published.
Mardia, K. V., Kent, J. T., Goodall, C. R. and Little, J. A. (1993) Kriging and splines with derivative information. *Technical Report STAT 93/15/C*. University of Leeds, Leeds.
Mardia, K. V., Kent, J. T. and Walder, A. N. (1991) Statistical shape models for image analysis. In *Computing Science and Statistics: Proc. 23rd Symp. Interface* (ed. E. M. Keramidas), pp. 550–557. Fairfax Station: Interface Foundation.
Mark, K., Miller, M. I., Grenander, U. and Abney, S. (1992) Parameter estimation for constrained context-free language models. *1992 DARPA Wkshp Speech and Natural Language, Feb.*
Miller, M. I., Joshi, S., Maffitt, D. R., McNally, J. G. and Grenander, U. (1993) Mitochondria, membranes and amoebae: 1, 2 and 3 dimensional shape models. *J. Appl. Statist.*, **20**, 137–159.
Miller, M. I. and O'Sullivan, J. A. (1992) Entropies and combinatorics of random branching processes and context-free languagues. *IEEE Trans. Inform. Theory*, **38**, no. 4.
Molina, R. and Ripley, B. D. (1989) Using spatial models as priors in astronomical image analysis. *J. Appl. Statist.*, **16**, 193–206.

Møller, J. (1989) On the rate of convergence of spatial birth-and-death processes. *Ann. Inst. Statist. Math.*, **41**, 565–581.
Mumford, D. and Shah, J. (1985) Boundary detection by minimizing functionals I. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. Los Alamitos: Institute of Electrical and Electronics Engineers Computer Science Press.
Parisi, G. (1981) Correlation functions and computer simulations. *Nucl. Phys.* B, **180**, 378–384.
Petrou, M. (1993) *International Mathematical Association Conf. Complex Systems and Engineering, Leeds, Sept.*
Phillips, D. B. and Smith, A. F. M. (1993) Bayesian faces via hierarchical template modelling. Submitted to *J. Am. Statist. Ass.*
Picard, R. W. (1991) Texture modeling: temperature effects on Markov/Gibbs random fields. *DSc Thesis*. Massachusetts Institute of Technology, Cambridge.
Preston, C. J. (1977) Spatial birth-and-death processes. *Bull. Int. Statist. Inst.*, **46**, 371–391.
Rao, C. R., Sastry, C. R. and Zhou, B. (1993) Tracking the direction of arrival of multiple moving targets. *IEEE Trans. Acoust. Spch Signal Process.*, to be published.
Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.
——(1981) *Spatial Statistics*. New York: Wiley.
——(1986) Statistics, images and pattern recognition. *Can. J. Statist.*, **14**, 83–111.
——(1988) *Statistical Inference for Spatial Processes*. Cambridge: Cambridge University Press.
——(1989) Gibbsian interaction models. In *Spatial Statistics: Past, Present and Future* (ed. D. A. Griffiths), pp. 1–19. New York: Image.
——(1991) The use of spatial models as image priors. *Lect. Notes Monogr.*, **20**, 309–340.
Ripley, B. D. and Kelly, F. P. (1977) Markov point processes. *J. Lond. Math. Soc.*, **15**, 188–192.
Ripley, B. D. and Sutherland, A. I. (1990) Finding spiral structures in images of galaxies. *Phil. Trans. R. Soc. Lond.* A, **332**, 477–485.
Sastry, C. R., Kamen, E. W. and Simaan, M. (1991) An efficient algorithm for tracking angles of arrival of moving targets. *IEEE Trans. Acoust. Spch Signal Process.*, **39**, 242–246.
Serra, J. (1982) *Image Analysis and Mathematical Morphology*. London: Academic Press.
Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L. and Cowell, R. G. (1993) Bayesian analysis in expert systems. *Statist. Sci.*, **8**, 219–247.
Stoyan, D., Kendall, W. S. and Mecke, J. (1987) *Stochastic Geometry and Its Applications*. Chichester: Wiley.
Sword, C. K., Simaan, M. and Kamen, E. W. (1990) Multiple target angle tracking using sensor array outputs. *IEEE Trans. Aerspce Electron. Syst.*, **26**, 367–373.
Timoshenko, S. and Woinowsky-Krieger, S. (1959) *Theory of Plates and Shells*, 2nd edn. New York: McGraw-Hill.